Exploring the effects of migration and admixture on human populations through time, using ancient DNA

Dissertation

zur Erlangung des akademischen Grades doctor rerum naturalium (Dr. Rer. Nat.)

Vorgelegt dem Rat der Fakultät für Biowissenschaften der

Friedrich-Schiller-Universität Jena

Von B.Sc. (with Hons) Genetics, M.Sc. Evolutionary Anthropology

Thiseas Christos Lamnidis

Geb. Am 23.06.1992 in Cholargos

Affiliation: Max Planck Institute for the Science of Human History, Jena

Date:09/02/2021

Declaration of Authorship

I hereby declare that I am the author of the present dissertation thesis.

_____

Gutachter:

1. Dr. Stephan Schiffels, Max-Planck-Institut für Menschheitsgeschichte

2. Prof. Dr. Holger Schielzeth, Friedrich-Schiller-Universität Jena

3. Prof. Dr. Anna Sapfo Malaspinas, Université de Lausanne

Beginn der Promotion: 20/04/2020

Dissertation eingereicht am: 19/07/2020

Tag der öffentlichen Verteidigung: 21/01/2021

Table of Contents

1. Introduction

Archaeogenetics is the research field of studying the genetic information contained in ancient DNA (aDNA) to gain insight into the past. The field emerged in 1984 with the publication of two studies using DNA recovered from archaeological material of an extinct equine species (*Equus quagga quagga*) (Higuchi et al., 1984), and human mummies from Egypt (Pääbo, 1984). These studies used bacterial cloning to amplify the ancient DNA, which was otherwise too sparse for direct sequencing. The subsequent invention of the polymerase chain reaction (PCR), which was able to target specific DNA sequences for amplification (Mullis & Faloona, 1989; Saiki et al., 1985, 1988), paved the way for amplification and sequencing of ancient DNA, while also overcoming biases that arose from the cloning process (Pääbo & Wilson, 1988).

However, the design of PCR, requiring two flanking 20 base-pair oligo-nucleotide amplification primers, limits amplification to DNA fragments exceeding 40 bp. The degraded nature of ancient DNA reduces the fragment lengths of ancient molecules considerably, thus limiting the ancient DNA yields of PCR approaches, while preferentially amplifying longer DNA fragments that are more likely to originate from present-day contaminants. Some ambitious early studies of ancient DNA that relied on this technology, like the analysis of 80-million-year-old DNA from Cretaceous period bone fragments (Woodward et al., 1994), were later revealed to have been the result of contamination (Hedges & Schweitzer, 1995).

In light of such findings, skepticism of aDNA studies grew in the scientific community. This skepticism regarding the authenticity of the purported ancient origin of the analysed DNA (especially for human aDNA) put significant focus on careful experimental design and authentication in the field (Handt et al., 1994; Hofreiter et al., 2001; Pääbo et al., 1989; Taylor, 1996), as well as sharing of published data (Anagnostou et al., 2015). With the invention of next-generation sequencing (NGS) in the last two decades (Margulies et al., 2005), the field has seen an expansion in interest from researchers as well as the public. In contrast to targeted PCR amplification and sequencing, NGS requires the preparation of “libraries” out of DNA molecules. These DNA libraries are made by ligation of sequencing adapters onto either end of a double-stranded DNA molecule. These libraries are then amplified via PCR, before sequencing. The ligation of the sequencing adapters artificially increases the length of fragments, thus ameliorating the amplification bias of PCR mentioned above. Additionally, the ability to sequence millions of DNA sequences in parallel reduces sequencing costs considerably, thus making sequencing of sparse aDNA more economically feasible.

This thesis will discuss the analysis of human aDNA, although it should be noted that archaeogenetic approaches have been successfully used in the past to study the population history of different animals (Der Sarkissian et al., 2015; Gretzinger et al., 2019; Orlando et al., 2002, 2008; Palkopoulou et al., 2018), aDNA of pathogens (Andrades Valtueña et al., 2017; Bos et al., 2016; Schuenemann et al., 2018; Vågene et al., 2018) and oral microbiomes recovered from dental calculus (I. M. Velsko et al., 2019; Warinner et al., 2014; Weyrich et al., 2017) retrieved from human remains, as well as environmental aDNA from the soil within caves (Slon et al., 2017).

Analysis of human aDNA from archaeological material has allowed archaeogeneticists to observe changes in the genetic composition of populations in an area through time. By using aDNA in this manner, a higher degree of resolution can be gained into the timing of past genetic

transitions, compared to the resolution that is available when inferring the past from modern genomic data alone. In recent years, technological advances in sampling of bioarchaeological material (Pinhasi et al., 2015), as well as enrichment of endogenous DNA against environmental contamination (Fu, Meyer, et al., 2013) have increased the amount of usable aDNA sequences retrieved for archaeogenetic analyses. These advances have allowed more thorough sampling of ancient populations, since they increase the proportion of sampled individuals that cross the minimum coverage thresholds required for comparative analyses. Sampling of ancient populations also allows direct observation of populations that may have become partly or completely replaced, or have become geographically restricted and hence harder to sample (Posth et al., 2018). Often, genetic transitions can be linked to changes in material culture, and can therefore be informative about the mode of transfer of cultural traits (Haak et al., 2015; Lazaridis et al., 2014, 2016). Archaeogenetic approaches have also been used to facilitate repatriation of remains without provenance, by identifying the strongest genetic affinities between the ancient individuals and sampled present-day populations (Wright et al., 2018), and to understand social organisation in ancient populations, by identifying genetic kinship between different individuals from the same burial ground (Mittnik et al., 2019).

However, archaeogenetics is but one of the disciplines utilised to study past human population history. While this approach can provide information that cannot be recovered by other archaeological or anthropological approaches (e.g. whether an individual shares a large proportion of their ancestry with sampled hunter-gatherers from the region), it is also blind to information that other disciplines can infer (e.g. whether that individual also had the diet/grave-goods of a hunter-gatherer). When attempting to reconstruct human population history it is therefore paramount that researchers take into account the evidence from different disciplines and engage in interdisciplinary discussion of their results.

In this thesis, I will focus on the movement of genes, via migration of people and/or admixture, and the information that this movement can provide about human history. I will introduce the differences between the inheritance mechanisms of uniparental (mitochondrial DNA and the Y-chromosome) and autosomal markers; the forces of evolution in population genetics; some methods commonly used in the analysis of human aDNA in the manuscripts included in this thesis; prior (archaeo-)genetics research regarding the population history of West Eurasia and the Americas - as context for my own research in these geographic areas - , and discuss the information gained by my own work about the population history of the areas studied, the limitations of archaeogenetic inferences, and the importance of combining archaeogenetic results with those from other disciplines when studying human history.

1.1 Introduction to human population genetics

Before discussing the methods used to analyse and model human population history using aDNA, I will introduce the inheritance patterns of the different parts of the human genome, and the four evolutionary forces, as well as briefly discuss how they apply to human population history.

1.1.1 Inheritance patterns in humans

Humans are eukaryotes, which means that the majority of the human genome is contained within the nucleus of each cell. Within each nucleus, human cells contain two copies of each of the 22 autosomes, and two sex chromosomes. The latter set of chromosomes determines the sex of each individual, with females having two copies of the X chromosome (XX), while males having one X and one Y chromosome (XY). In addition to nuclear DNA, human cells also carry multiple copies of a mitochondrial genome (mtDNA) that is 16,569 base pairs long (Anderson et al., 1981).

The human mtDNA and most of the Y chromosome (Y-chr) are non-recombining genetic elements and thus inherited in their entirety from the maternal and paternal lineages respectively. The lack of recombination in these genetic elements causes them to be inherited unchanged from one generation to the next, with the exception of *de-novo* mutations. Therefore, the variation of these genetic markers can be captured within a phylogeny, without requiring any horizontal transfer between the branches. Adding an assumption on the mutation rate or the split time of certain lineages, based on previous research, to these phylogenies enables the estimation of a mutation rate (when this has not been assumed) and/or estimated split times for each lineage split (Friedlaender et al., 2005; Fu, Mittnik, et al., 2013; Soares et al., 2009; Wei et al., 2013).

The human autosomes and X chromosome, which make up the majority of the human genome, undergo recombination. Recombination takes place during gamete generation, when the chromosomes are arranged in pairs during the interphase of meiosis. In this arrangement, the paternal and maternal chromosomes can cross over, effectively exchanging a section of the DNA between them. As a result, the majority of an individual's genome is inherited as a patchwork of sections from the genomes of their ancestors. Over time, recombination events will shorten the stretches of DNA that are inherited together (i.e. haplotypes), thus decreasing the effects of genetic linkage.

Studies on population history have often analysed mtDNA and Y-chr haplotypes to identify relationships between human populations (Ilumäe et al., 2016; Kivisild, 2015; Nesheva, 2014; Underhill et al., 2000), and these approaches are often used as unique tools for the identification of female- and male-specific aspects of human population history (Mittnik et al., 2019; Saag et al., 2017). Additionally, the large copy number of mtDNA increases the likelihood of its preservation in archaeological material and decreases the likelihood of contamination with modern DNA, features which made many earlier archaeogenetic studies focus on mtDNA (Green et al., 2008; Haak et al., 2005). However, analysis of autosomal DNA provides a more complete picture of a population's history, because each sampled individual reveals genetic information inherited from multiple ancestors. The aforementioned patchwork of genetic information can reveal even low levels of admixture (especially from distant sources), even when these would not be otherwise detectable. A prominent example of this is the revelation of low levels of gene-flow from Neanderthals and Denisovans into different human populations, that was only detected when analysing the nuclear genomes of these hominins (Green et al., 2010; Meyer et al., 2012). Additionally, by estimating haplotype lengths on the recombining sections of the human genome closely following a selective sweep or admixture event, it is possible to estimate the timing of these events.

1.1.2 The forces of evolution

The allele frequencies of a population can change as a result of four evolutionary forces: mutation; gene-flow; selection; and genetic drift. Mutations take place when errors that happen during copying of a DNA molecule introduce permanent changes to the sequence of the resulting DNA copy. Mutations can also occur due to environmental factors like exposure to UV light (Ikehata & Ono, 2011). Gene-flow refers to the introduction of genetic variation from another population as a result of admixture. Following an admixture event, the allele frequencies of the resulting population will be intermediate between the allele frequencies of the source populations. Together, these processes increase the genetic variation in a population.

Selection acts on existing genetic variation within a population, directionally increasing and/or decreasing allele frequencies in variants. There are many types of selection, each with distinct effects on the allele frequency of the affected variants. Some types of selection will maintain genetic variation on a specific variant (e.g. stabilising selection on the causative variant of sickle-cell anaemia (Bunn, 2013)), while other types will eventually drive an allele to fixation, thus reducing genetic variation in the population. The human genome is considered to be largely devoid of selection. Specifically, a maximum of 15% of the human genome is expected to be functional (Graur, 2017), while roughly 10% or less is estimated to be under the effects of selection (Rands et al., 2014).

Genetic drift refers to the change in the frequency of an existing allele in a population resulting from the randomness in the reproductive success of individuals in the parent generation. This change is not directional, hence the expected allele frequency for each generation stays constant. The effects of genetic drift are more pronounced in smaller populations, and slowly become almost negligible as the size of a population grows.

1.2 Analytical methods used in ancient DNA

Before discussing the specific analytical methods used in ancient DNA, it is important to state two of the assumptions that are commonly, and often implicitly, made in population genetics studies. First, the use of neutral models is common when studying human history, which assume the lack of any selection across the human genome, an assumption which is empirically validated, as mentioned above. Second, is the assumption that *de novo* mutations will not occur in the same position of the genome more than once. This assumption is unproblematic in relatively small sample sizes because of the low per-base mutation rate observed in humans.

A direct consequence of this assumption is that any shared variation between two populations must share a common ancestor, and did not arise from two independent mutation events. This consequence is of utmost importance for studying human population history. Multi-allelic sites (which would require a minimum of two mutations to have occurred on the same site for tri-allelic sites) are rare in the sample sizes of most studies of human history and are normally excluded from any analyses of human population history. In fact, the extremely low number of sites removed by such filtering adds credence to the assumption that no secondary mutations take place. However, this filtering does not account for secondary mutations (i.e. back-mutations and independent mutations). These mutations will result in the same allele

being present in multiple populations without reflecting a shared population history. Luckily, such events are rare in human populations. Given the low per-base-per-generation mutation rate ($\mu \approx 1.1\times10^{-8}$) (Narasimhan et al., 2017) the probability that a specific allele is shared between two individuals due to an independent mutation instead of shared population history is $p = \frac{1}{3}\times\mu\times2n \approx 1.5\times10^{-8}$, where μ is the per-base-per-generation mutation rate, n is the number of diploid individuals, and $\frac{1}{3}$ reflects the probability that both mutation events resulted in the same nucleotide misincorporation. Although this probability is small, as the number of studied individuals rises, so does the probability of observing secondary mutations. Indeed, if we take into account the census size of the human population (7.8×10^9 as of March 2020 (*World Population Clock: 7.8 Billion People (2020) - Worldometer*, n.d.)), we can estimate that each generation, at any given site, roughly 172 people carry an allele that is the result of a *de novo* mutation. However, the probability of observing a secondary mutation remains below 0.001 for sample sizes up to 100,000 individuals, well above the typical number of individuals analysed in studies of human population history. Such mutations are thus unlikely to affect the results of genetic analyses of population history significantly, especially when many sites along the genome are analysed in concert thus lowering the weight of any single site.

From the equation above, and the size of the human population in the present, it becomes evident that any position across the human genome will have undergone mutation in some individuals in each generation. It follows then that every site along the human genome will have segregating polymorphisms, except those sites where a mutation would be embryonic lethal. Since the majority of the human genomes is evolving neutrally, the allele frequency for these variants is dominated by drift and gene-flow/admixture. Below, I will outline two approaches to studying human history through modelling the effects of drift and admixture events on common and rare alleles respectively.

1.2.1 Analysis of common variants in human population history

Many studies analysing human aDNA resort to limiting their analysis to a subset of single nucleotide polymorphisms (SNPs) known to be polymorphic within human populations. In order to effectively model human history on subsets of SNPs, it is important to know the exact ascertainment of the SNP panel being used. There are two such SNP panels that are commonly used in ancient DNA studies. The first is the Affymetrix Human Origins array. This SNP panel contains SNPs that were observed as heterozygous in a single individual from various worldwide human populations, as well as SNPs where a randomly chosen San individual carried a derived allele relative to sequencing data available from archaic hominins (Reich et al., 2012). After multiple quality control filters and the addition of SNPs on the Y-chr, mtDNA and other commercially available arrays, this dataset contains roughly 600,000 SNPs. The second dataset is a supergroup of the Human Origins array that includes roughly 600,000 additional SNPs from the Illumina 610-Quad array, complemented with additional SNPs from the X- and Y-chr, as well as roughly 47,000 SNPs of functional importance (Mathieson et al., 2015). There are multiple advantages to limiting genomic analysis to a specific SNP panel. On one hand, the cost of genotyping present-day individuals on such a panel is considerably cheaper, and therefore a

larger set of present-day comparative data is available for these markers compared to entire genomes. On the other hand, with the development of in-solution capture protocols (Fu, Meyer, et al., 2013) (which enable the enrichment of human DNA overlapping specific SNPs from a genomic library), it is possible to analyse individuals with poor DNA preservation, while also making the generation of plentiful data from individuals with good DNA preservation more economically feasible.

The most popular methods for analysis of such SNP data are F- and D-statistics (Patterson et al., 2012; Reich et al., 2009). These offer a way to quantitatively test specific phylogenetic hypotheses and estimate the amount of shared population history between two individuals/populations. Additionally, each of these statistics comes with a standard error, calculated via block jackknifing which counteracts the effects of linkage between SNPs and can be used to calculate the degree of statistical significance of some of these statistics. The simplest form of an F-statistic is the F_2 statistic. F_2 between populations A and B is defined as the average across all sites of the square of allele frequency difference between populations A and B. For a single site, this is defined as: $F_2(A,B) = (a' - b')^2$, where a' and b' are the allele frequencies of populations A and B respectively. This statistic corresponds to the branch length between the two tested populations measured in units of "drift-time" (see Figure A below). This compound unit of time and population size is a consequence of the fact that drift does not change the allele frequency of a variant (i.e. "accumulate") along a phylogenetic branch at a constant rate, but instead as a function of the size of the population, and hence the two cannot be disentangled using these statistics.

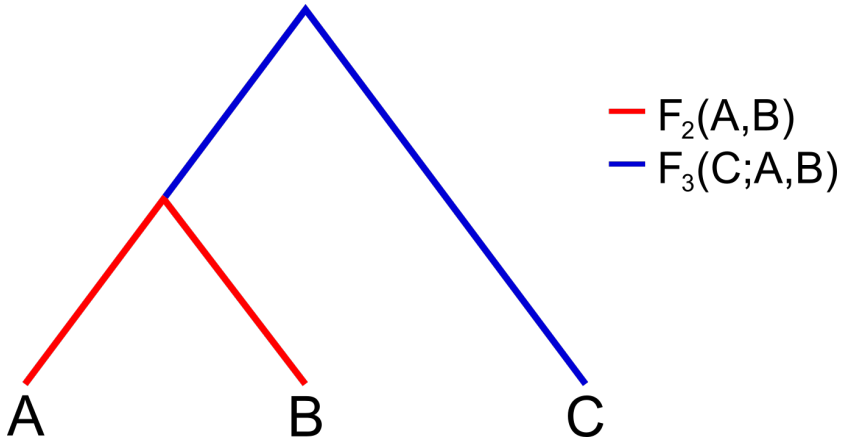


Figure A. A visual representation of the branch lengths measured by F_2 and F_3 statistics on a simple phylogeny of three populations. *F_2(A, B)* shown in red. *F_3(C; A, B)* shown in blue.

A more commonly used statistic to infer population history if the F_3 statistic. This is an extension of the F_2 that is defined as the average across all sites of the product of the allele frequency differences of populations A and B to population C. For a single site, this is defined as $F_3(C;A,B) = (c' - a')(c' - b')$, following the same notation as above. This statistic has two forms which can be used to a) test if a population C can be explained as the result of admixture between populations A and B (i.e. it has an intermediate allele frequency), and b) estimate the amount of drift populations A and B share from an outgroup population C. In the first form, if c' is intermediate to a' and b', the value of the F_3 above becomes negative. Combined with the Z score of the calculated F_3 value, this acts as a statistical test that population C is the result of

admixture from two source populations that are related to populations A and B. In the latter form, the statistic is constructed so that population C is a distant outgroup to populations A and B. The statistic then measures the amount of drift that has accumulated between populations A/B and C, and is shared between populations A/B. On a population phylogeny, this measure corresponds to the branch length from population C to the common ancestor of populations A and B (see Figure A above). When using F3 statistics of this form, it is important to ensure that population C is indeed an outgroup population that is symmetrically related to populations A and B. For this reason, it is common to use the Chimpanzee reference genome in that position when studying the interactions between hominin species, while studies of human populations outside of Africa will often use Sub-Saharan African populations in this position (e.g. Mbuti, a hunter-gatherer group indigenous to the Congo, or Yoruba, a western African population).

The final set of statistics I will be introducing here are the D and F_4 statistics. These statistics are very similar, but normalised in a slightly different way. When considering a phylogeny of four individuals (W, X, Y, Z) like the one shown in Figure B below, we define $D(W,X;Y,Z) = \frac{nBABA - nABBA}{nBABA + nABBA}$. In this notation, ABBA denotes a sharing pattern where individuals W and Y have concordant alleles, as do individuals X and Z. BABA is defined with the same notation, and n denotes the number of sites with the sharing patterns in question. A D-statistic can be intuitively thought of as a test for "treeness", or the degree to which the observed data deviate from a specified tree. Extending the definition of D-statistics to populations and using allele frequencies instead we get $D(W,X;Y,Z) = \frac{\sum_i (w'_i - x'_i)(y'_i - z'_i)}{\sum_i (w'_i + x'_i - 2w'_i x'_i)(y'_i + z'_i - 2y'_i z'_i)}$, where w'_i is the allele frequency of population W at SNP i, and the numerator and denominator are each summed over all sites. If the specified tree-phylogeny is correct and no admixture has taken place between the tested populations, the expected value of a D-statistic is 0. In cases where the statistic is not equal to 0, the allele sharing between the four tested populations does not match the expected sharing given the specified phylogeny. The polarity of the statistic will then reveal if there is an excess of ABBA or BABA sites, and hence in what way the specified phylogeny is violated. When *D(W, X; Y, Z)* is significantly positive, it can be interpreted as gene-flow between populations W and Y, or X and Z, either of which breaks the "treeness" of the data. In order to distinguish between these two scenarios, is it customary to include one outgroup population in such statistics. For example, the increased number of ABBA sites reported by Green et al. (2010) with the use of $D(African, European, Neandertal, chimpanzee)$[1], could theoretically be caused by excess allele sharing between Chimpanzee and Africans, or between Europeans and Neanderthals. However, because Chimpanzee is a true outgroup population, we can discount that any attraction between Chimpanzee and Africans exists, and conclude that the value of the D-statistic is caused by gene-flow between Neanderthals and the ancestors of Europeans.

In similar notation to that used above, we can define $F_4(W,X;Y,Z) = \frac{\sum_i (w'_i - x'_i)(y'_i - z'_i)}{\sum_i 1}$. The numerator remains the same as that of D-statistics, but the denominator in this case is

[1] Some implementations of F_4 and D-statistics use *nABBA - nBABA* instead. This results in a reversed polarity in the resulting values of these statistics to the one described here. Indeed Green et al. (2010) refer to a positive value for this statistic, even though according to the definitions given here it should result in a negative statistic. As long as one is aware of which implementation of the statistic they are using, the interpretation of the values remains consistent.

equal to the number of sites used in the calculation and not limited to just the ABBA/BABA sites. This statistic behaves similarly to D-statistics, having an expected value of 0 when the assumed specified phylogeny is correct and no admixture has taken place, and otherwise having a polarity specifying which populations appear more closely related than they are expected to be if the specified phylogeny is correct. However, much like F_2 and F_3 statistics, the value of an F_4 can be interpreted as a branch length in a population phylogeny. Specifically, an F_4 statistic corresponds to the branch length of the internal branches of a phylogeny, weighted by the admixture proportions leading to those edges (see F_4 (V, W; X, Y) highlighted in Figure B). Because the statistical significance of an F_4 as a test for admixture depends on the statistical uncertainty (typically assessed using a bootstrapping approach) of the measurement, the resolution of this approach is limited when the allele frequency difference between the source populations of an admixture event is small, and hence the internal branches are short, or when the admixture proportion from a source population to the admixed population is small (V, X and β in Figure B respectively), and hence the proportion of alleles that have accumulated drift along the internal branch is smaller. In both of these cases, it becomes difficult to statistically distinguish the value of the F_4 from 0, and hence large numbers of individuals will need to be sampled from a population before an admixture event can be identified. Nevertheless, these statistics were shown to be able to detect gene flow of down to a few percent in case of relatively diverged source populations, such as those studied here.

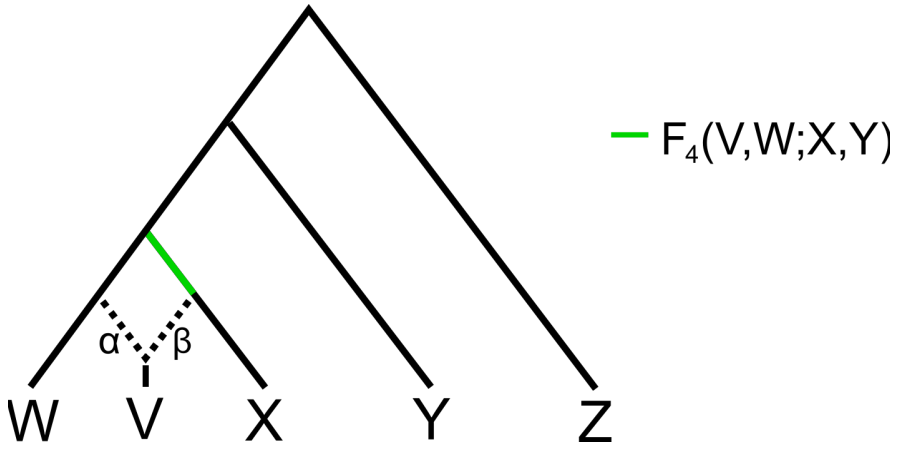


Figure B. A simple phylogeny of 5 populations. Population V is the result of an admixture event between two populations related to populations W and X. The branch length measured by F_4(V,W;X,Y) is highlighted in green and is proportional to β.

1.2.1 Using rare variation to study human history

An alternative approach to the study of common variation across human populations is to use rare variation to build demographic models of human populations history and study the interaction between different human populations. I use this approach in Manuscript B, as implemented in *rarecoal*, which I will be introducing below.

Rarecoal is an approach of computing the joint allele frequency spectrum for rare alleles given an arbitrary demographic model connecting the populations tested. The demographic model can include population splits, population size changes (e.g. bottlenecks), and single-burst

admixture/migration events of different proportions. This computation is done on the framework of a neutral coalescent model, and therefore all events are described from a perspective of looking backwards in time from the present. In such a framework, different alleles coalesce at a rate that is inversely proportional to the size of the population in the previous generation. According to standard neutral coalescence theory, the rate at which lineages coalesce is equal to $\frac{1}{2N}$, where N is the population size of the population. This can be thought of as the probability that two lineages originate from the same ancestor in the previous generation. When looking backwards through time, the number of coalescence events per unit time decreases, because only a fraction of the alleles within the population survive each successive generation. In fact, the amount of time between each successive coalescence event increases exponentially, meaning that most coalesce events take place in the recent past.

This knowledge creates a certain expectation when looking at sharing patterns of rare alleles between different populations, namely that more rare alleles will be shared between populations that have split more recently, than between populations with a deeper split. For example, we expect that a European individual will share more rare alleles with another European individual than with an East Asian individual, who in turn will share more rare alleles with another individual from East Asia than with either of the two European individuals. A case where this expectation might break down, is when admixture has occurred. In such cases, lineages from the admixture source population have been introduced into the target population at the time of admixture (when looking forwards in time), effectively overwriting a proportion of the lineages within the target population equal to the admixture proportion. As expected, this increases the rare allele sharing between the admixture target and source populations. Population splits between two populations are modelled in a similar way, except all the lineages of one population are merged entirely into the other (in the backwards time perspective). Within *rarecoal*, each demographic event is specified with a set of parameters, like the timing of a population split or an admixture event, the admixture proportion, the population size of a branch etc. *rarecoal* optimises the parameters of the specified demographic events to minimise the differences between the observed rare allele sharing patterns and the expected computed joint rare allele frequency spectra with the use of maximum likelihood or Markov Chain Monte Carlo approaches (with *rarecoal maxl* and *rarecoal mcmc* respectively). This results in a demographic model with inferred model parameters, each with its own confidence interval, as well as a likelihood score that can be used to compare different models.

An important consideration for this approach is the definition of a rare allele, as well as why rare alleles are used specifically. In this context, an allele is defined as rare depending on the number of times it is observed in the set of present-day genomes used for demographic modelling. This cutoff is therefore defined based on the allele count, and not the allele frequency of the variant (although the two are related), meaning that as more populations are added to a demographic model, and the allele count cutoff is kept constant, the allele frequency cutoff for a rare allele becomes smaller. The reason for using specifically rare alleles, and for defining them based on their allele count and not their allele frequency, has to do with the computational intensity of this approach for larger allele counts. Alleles that have been observed only c times in a set of p populations can be arranged in multiple different patterns of allele sharing. The number of unique patterns of sharing that can be observed is equal to $\frac{(p+c-1)!}{p!(c-1)!}$. When defining a

maximum allele count of 4 in a demographic model of 5 populations, the observed data comprises all sharing patterns for allele counts 1 to 4, totalling 125 different unique sharing patterns. Increasing the maximum allele count to 5 would more than double the numbers of observations, thus significantly increasing computational intensity. Beyond increasing the number of unique sharing patterns across populations in the present, increasing the maximum allele counts will increase the complexity of the coalescence process as that is traced backwards in time. This added complexity is the result of all the possible sharing patterns that can arise from each single sharing pattern in the present after each coalescence event has occurred, a matrix of possibilities that grows considerably more complex with each increase to the allele counts considered.

1.3 The major genetic transitions in Western Eurasia from the Mesolithic to the Bronze Age.

West Eurasia is, thus far, the region with the highest concentration of analysed ancient genomes, both spatially and temporally. In part, this sampling density is a result of the climatic conditions in this geographical area, which allow for better preservation of DNA in archaeological material (Kistler et al., 2017). As a result, no better example can be given currently about the information that archaeogenetics can reveal than the population history of West Eurasia from the Mesolithic to the Bronze Age. Between the Mesolithic and the Bronze age two major genetic transitions occurred in West Eurasia, one during the Mesolithic to Neolithic transition, and one during the Bronze Age.

The Mesolithic to Neolithic transition in Europe entails the shift from hunter-gatherer societies to societies that utilise agriculture and are characterised by a sedentary lifestyle, pottery and domesticated animals. This technological shift began 12,000-10,000 years before present (yBP) in the Levant, and spread to Europe gradually between 9,000 and 3,500 yBP (Barker, 2006; Isern et al., 2012). The mode by which the Neolithic revolution was introduced to Europe has been under debate among archaeologists, with two opposing demographic scenarios: demic diffusion vs cultural diffusion. Shortly, the demic diffusion scenario argues that the Neolithic package was brought to Europe through large scale movements of people from the Levant who migrated to Europe, while the cultural diffusion scenario claims that hunter-gatherer societies adopted agriculture progressively (Svizzero, 2017). Analyses of genomes from multiple ancient individuals from the Neolithic period in different parts of Europe and the Near East have shown that the Neolithic transition in Europe corresponds to a large genetic shift, consistent with a demic diffusion from the Near East (Haak et al., 2010; Hofmanová et al., 2016; Kılınç et al., 2016; Lazaridis et al., 2014, 2016; Mathieson et al., 2015, 2018; Olalde et al., 2015). The genetic evidence supports a single source for Anatolian ancestry in most of Europe, with the notable exception of southern Greece (Mathieson et al., 2018). However, following this initial movement of early farmers into Europe, groups of hunter-gatherers and farmers began admixing, giving rise to populations of dual ancestry (Haak et al., 2015; Lipson et al., 2017; Mathieson et al., 2018), a result which supports some degree of cultural transmission albeit at a later stage.

About 5,300 yBP, the Yamnaya culture formed in the western Eurasian steppe (Morgunova & Khokhlova, 2013). This population was found to be a mixture of two distinct West Eurasian genetic components (Haak et al., 2015); one related to local hunter-gatherers who showed a genetic affinity to a 24,000 yBP Siberian individual (Raghavan et al., 2014), and another related to hunter-gatherers from the Caucasus (Jones et al., 2015). By 4,500 yBP, ancestry related to the Yamnaya population was detected in individuals from Germany associated with the Corded Ware culture. Three quarters of the ancestry in the Corded-Ware-associated population was attributed to the Yamnaya-related component, while the remainder was attributed to ancestries predominantly identified in hunter-gatherer and early Farmer groups (Haak et al., 2015). This result showcases a population movement that brought this ancestry into Central Europe, but also admixture between the local dual-ancestry populations and the incoming Yamnaya-related population. It has been suggested that this population movement is responsible for the spread of Indo-European languages into Europe (Haak et al., 2015), although this hypothesis is seen as contentious among linguists.

Haak and colleagues (2015) successfully modelled most present-day European populations as mixtures of the three genetic components discussed so far. Notably, four present-day populations were identified that did not fit a three-source model, namely: Maltese, Mordovians, Finns, and Russians (Haak et al., 2015). The authors found that adding a fourth source population from Siberia, the Nganasan from the Taymyr peninsula, produced well fitting models for Mordovians, Finns and Russians. This result showcases a genetic connection between present-day populations from northeastern Europe and those from East Eurasia, that has been previously identified (Salmela et al., 2008).

In Manuscript A, we set out to better describe this genetic connection. In line with Haak and colleagues (2015), we identify this genetic component as being most closely related to present-day Siberian populations, with the Nganasan population being the best proxy for the original source of this ancestry in Fennoscandia. We report the earliest direct evidence for the presence of Siberian-related ancestry in the region, in a population that we identified via ancient DNA from the island of Bolshoy Oleni Ostrov dated to 3,500 yBP. We estimate that this ancestry component first entered the Bolshoy population roughly 17 generations (equivalent to about 500 years) before the time of the analysed individuals. Additionally, through analysis of ancient and present-day genomes covering the last 3,500 years, we identify a dilution of Siberian-related ancestry in the area through time, and note that Uralic-speakers tend to have higher proportions of Siberian ancestry than their non-Uralic-speaking neighbours. Finally, genetic analysis of Iron Age remains from a water burial in Southern Ostrobothnia dated to 300–800 CE, revealed that this population was genetically more similar to present-day and historical (100 yBP) Saami populations than to present-day Finns.

1.4 The population history of the Americas

The American supercontinent was populated in multiple successive waves of people that are thought to have entered through the Bering land bridge. It has been suggested that the first people to enter the supercontinent arrived through the southern coast of the Bering land bridge, and not the ice-free corridor between the Cordilleran and the Laurentide ice sheets, although

the latter may have been used by later migrations into the supercontinent (Pedersen et al., 2016). Genomic analysis of present-day Native American populations has revealed that all present-day Native American populations are descendants of three ancestral lineages (Reich et al., 2012). The first lineage has been modelled as having dual ancestry, made up of a West-Eurasian lineage identified through ancient DNA from a 24,000 year old Siberian individual, and a genetic lineage identified in present-day East Asians (Raghavan et al., 2014). This is the earliest population to enter the Americas, while still contributing to the present-day gene pool of the supercontinent, and is thought to have split from other Siberian populations 16,000-22,000 years ago (Moreno-Mayar, Potter, et al., 2018; Raghavan et al., 2014). Within this lineage, a population separation took place roughly 16,000 yBP between the ancestors of present-day North and South Americans (Moreno-Mayar, Potter, et al., 2018). The second and third ancestry fluxes into the Americas happened more recently and are only observed in some North American populations. They are associated with populations speaking languages of the Na-Dene and Eskimo-Aleutian families respectively (Reich et al., 2012). Each of these two lineages shares part of its ancestry with an ancient individual related to the first known culture to settle Greenland (Flegontov et al., 2019; Rasmussen et al., 2010; Reich et al., 2012).

Although present-day Native South American populations derive their ancestry only from the first lineage described above, the population history of South America remains considerably complex. Recent studies analysing aDNA report evidence of multiple waves of migration between North and South America (Moreno-Mayar, Vinner, et al., 2018; Posth et al., 2018; Scheib et al., 2018). Posth et al. (2018) infer at least three ancestry fluxes into South America: one ancestry related to an ancient individual associated with the Clovis culture and dated to 12,500 yBP (Rasmussen et al., 2014), which was later replaced; one ancestry that was geographically restricted to the Central Andes and is related to a sampled population from the California Channel islands dated to roughly 5,500 yBP (Scheib et al., 2018); a final ancestry line that is prevalent among present-day Native South American populations. Analysis of genome-wide markers in populations from the Amazon basin revealed an additional ancestral component in the Karitiana and Surui populations, related to Andaman Islanders from South Asia (Skoglund et al., 2015). This finding, however, is hard to explain and has been characterised as a possible false positive result by one study that failed to detect any similar genetic signals in any sampled ancient population from South America in the study (Posth et al., 2018).

During the 2nd millennium current era (CE), contact between the Old World and the Americas altered the genetic landscape of the Americas significantly. The colonial era genocide on formerly large indigenous populations in the Americas (Dunbar-Ortiz, 2014; Stannard, 1993) followed by/alongside the introduction of diseases from the Old World to the supercontinent created a significant bottleneck in Native American populations (O'Fallon & Fehren-Schmitz, 2011), while admixture with populations that migrated from the Old World to the supercontinent (as part of colonisation and/or the slave trade) introduced detectable and variable proportions of ancestry from European and African lineages in many Native American populations (1000 Genomes Project Consortium et al., 2015; Montinaro et al., 2015). To counteract the effects these demographic events can have on pre-contact population history inferences, geneticists working with present-day populations from the Americas will often mask the genomes of individuals that show signals of admixture from European/African populations (Reich et al.,

2012), limit their datasets to individuals with little to no Old-World ancestry (Lazaridis et al., 2014; Mallick et al., 2016; Scheib et al., 2018), or explicitly model these events (Flegontov et al., 2019). Co-analysing ancient populations predating European contact and present-day populations in areas with genetic continuity can identify signatures of adaptation, while circumventing the confounding factors brought about by contact with the Old World (Lindo et al., 2018).

Manuscript B focuses on the population history of Fuego-Patagonian populations in the last 5,000 years, and how these populations are related to other present-day and ancient Native American populations. We analyse the genomes of 15 ancient individuals together with previously published ancient and present-day individuals from Tierra del Fuego, and the rest of South America. This dataset spans a temporal range of 4,500 years. To better detect genetic shifts through time within each group of ancient individuals sharing a population attribution, we employed an analytical approach that estimates the number of ancestry lines required to explain each group of ancient individuals sharing a population attribution. Individuals attributed to each Fuego-Patagonian population were grouped in the minimum number of genetically homogeneous subgroups per population using *qpWave (Reich et al., 2012)*. The individuals in each resulting subgroup could be modelled as being symmetrically related to a set of present-day Native American populations, while individuals from different subgroups are not. The resulting subgroups were then added to a previously-published (Posth et al., 2018) admixture graph model which fits the observed genetic variation within South America well, to investigate the relationship of Fuego-Patagonian populations to ancient populations in the rest of the continent. We observe that ancient Fuego-Patagonian groups derive the majority of their ancestry from a lineage most closely related to a sampled population from the Los Rieles archaeological site in Chile, and dated to 5,100 yBP. However, in order to successfully model Fuego-Patagonian groups from the last 1,100 years an additional minor contribution from a lineage related to an ancient population that inhabited the California channel islands circa 5,200 yBP was required. I will refer to this additional contribution as the Channel-Island-related ancestry/component from here on, although it does not imply a migration of people directly from the California Channel Islands into Fuego-Patagonia, but of a population that appears more closely related to the Channel Islands population than any other ancient individuals. Additionally, we created an admixture graph model of present-day Native American populations, to which we added each Fuego-Patagonian subgroup in an attempt to characterise their genetic relationships to present-day Native American populations. We observe that groups lacking the Channel-Island-related component can be modelled as most closely related to the Quechua population, while those groups with this ancestry component are best modelled as more closely related to Amazonian populations (Surui & Piapoco). Using the rare variation in a dataset of present-day individuals, we build a demographic model that fits the observed data well, estimate population split times, and infer the timing and strength of admixture events between these reference populations. The inferred population split between Europeans and East Asians is roughly 36 kya, while that of the ancestor of Siberian and Native American populations from East Asian populations is roughly 23 kya. To better infer the population structure of the Siberian/Native American ancestral population, we modelled the West Eurasian contribution into these lineages (Raghavan et al., 2014) as two admixture events, one into the ancestral population and one into the Native American population, occurring on concurrently with the split

of Siberian and Native American populations, roughly 17 kya. Within the Native American populations, first the Pima split from the Quechua/Piapoco ancestral lineage about 10 kya, with the remaining lineages separating 8 kya. We then use the rare variation that each ancient individual in our dataset (with mean genomic coverage above 1x) and the present-day reference data share to infer the most likely branching point for these ancient individuals on the aforementioned demographic model. Our results show that ancient Fuegian individuals share an equidistant relationship to present-day South American populations, within our resolution. Finally, multiple lines of evidence reveal no signs of a genetic contribution to ancient Fuego-Patagonians from European or African populations.

Staying thematically within the Americas, in Manuscript C, an interdisciplinary analysis of three individuals from a mass grave in Mexico City is conducted to identify them as first-generation Africans in Mexico, to reconstruct their life histories, and to gain insights into the cause of their deaths and the effects of the transatlantic slave trade on disease dissemination across the world. Dental modifications on these individuals hinted at an African origin for these individuals. We analysed the autosomal, immunogenetic, and uniparental markers of these individuals to investigate this hypothesis. All three individuals are most closely related to present-day Sub-Saharan and particularly Western African populations, and bear immunogenetic and uniparental haplotypes as well as skeletal features found among the present day variation of populations from these areas. Additionally, analysis of the autosomal markers of these individuals revealed no genetic contributions from non-African populations in these individuals. Furthermore, radiogenic isotope analysis revealed signatures consistent with Western African origin. Prompted by the burial of these individuals on the grounds of a hospital we screened the genetic data for possible pathogenic agents. Genetic signatures of a Hepatitis B virus (HBV) infection, and an infection by *Treponema pallidum subsp. pertenue* were identified in two individuals respectively. Genetic analysis of the HBV strain grouped it with present-day strains from The Gambia and Guinea. The analyzed treponemal genome was found to be most closely related to another ancient *Treponema* genome isolated from human remains from colonial Mexico (Schuenemann et al., 2018), while both of these ancient strains are associated with strains isolated in present-day Ghana.

2. Aim of the thesis

The aim of this thesis is to investigate the information archaeogenetics can reveal about human history by studying the movement of genes through time, as a result of population movements and admixture between populations. The manuscripts detailed in this thesis present three separate genetic movements in different spatio-temporal contexts: A) The origin and spread of Siberian ancestry in northeastern Europe in the last 3,500 years, as well as the extent of Saami-like ancestry in Finland since the Iron Age; B) The population dynamics of Fuego-Patagonian populations in the last 4,500 years, with a specific focus on the spread of an ancestry component maximised in a population from the island of San Nicolas in California in Fuego-Patagonia; and C) The identification of the first genetically-confirmed African individuals in the New World, where African ancestry is now observed in many Native American populations, and the reconstruction of their life histories using interdisciplinary analyses.

More specifically, the aims and questions addressed by these manuscripts and this thesis as a whole are:

Manuscript A:

- What is the extent of Siberian ancestry in present-day and ancient northeastern European populations?
- When did this ancestry first arrive in the area?
- Which present-day population do the Iron Age inhabitants of Finland most resemble genetically?

Manuscript B:

- How do ancient populations from Tierra del Fuego relate to present-day South American populations?
- Which ancestry wave into South America are Fuego-Patagonians derived from?
- Are there any genetic shifts in Fuego-Patagonian populations through time? When did these shifts take place?

Manuscript C:

- Showcasing the breadth and depth of information that is obtainable about the life and death of individuals via interdisciplinary archaeogenetic, osteological, archaeological and ethnological research.
- Can we use combined archaeogenetic and strontium isotope results to identify these individuals as first-generation Africans?
- Were these individuals slaves, brought to Mexico because of the slave-trade?
- Given their burial in a hospital, can we identify pathogens that may have caused the death of these individuals? Do the identified pathogens match well with the weaknesses of the immunogenetic haplotypes of the individuals?

- Is the Hepatitis B virus (HBV) genotype identified in one of these individuals consistent with directional movement of HBV between Africa and the Americas?

3. Overview of manuscripts and author's contribution

3.1 Manuscript A

"Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe"

Thiseas C. Lamnidis, Kerttu Majander, Choongwon Jeong, Elina Salmela, Anna Wessman, Vyacheslav Moiseyev, Valery Khartanovich, Oleg Balanovsky, Matthias Ongyerth, Antje Weihmann, Antti Sajantila, Janet Kelso, Svante Pääbo, Päivi Onkamo, Wolfgang Haak, Johannes Krause, Stephan Schiffels

Published in *Nature Communications*, 27 November 2018
Lamnidis, T.C., Majander, K. et al. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. Nat Commun 9, 5018 (2018). https://doi.org/10.1038/s41467-018-07483-5

In this manuscript, we report genome sequences for 15 ancient individuals from Fennoscandia with a temporal range from 1,500 BCE to the 19th century CE. In addition, we released additional sequencing data from one previously-published present-day Saami individual. We show that the genetic makeup of northeastern Europe was shaped by genetic contacts with Siberia that began at least 3,500 years ago. We find evidence of this Siberian ancestry, best proxied by the Nganasan population of the Taymyr peninsula, in present-day northeastern European populations, especially those speaking Uralic languages. We estimate that Siberian ancestry was introduced in the oldest sampled population from the Russian Kola peninsula roughly 4,000 yBP. Finally, we show that Iron-Age individuals from Southern Ostrobothnia in Finland are genetically closer to present-day and historical Saami than to Finnish populations, confirming the claim that the ancestors of the Saami had a broader geographic range within Finland in the past than today.

Author contributions:

- These authors contributed equally to this work: Thiseas C. Lamnidis, Kerttu Majander.
- These authors jointly supervised this work: Päivi Onkamo, Wolfgang Haak, Johannes Krause, Stephan Schiffels.

The design of this study, as well as collection of archaeological material for sampling were done prior to my involvement in the project by Anna Wessman, Vyachaslav Moiseyev, Valery Khartanovich, Antti Sajantila, Päivi Onkmo, Oleg Balanovsky, Wolfgang Haak and Johannes Krause. The laboratory procedures for this study, which were undertaken by Kerttu Majander, were comprised of sampling of archaeological material, DNA extraction, and library preparation. The enrichment of libraries for human DNA, as well as sequencing of the enriched libraries was done by the technical staff of the Max Planck Institute for the Science of Human History.
I was provided with raw sequencing data from 15 ancient individuals. I mapped the raw sequencing data to the human reference genome, verified the ancient origin of the DNA by quantifying the deamination patterns, pulled down the reads overlapping 1.2 million variable

positions and generated genotypes calls at these positions. I merged these genotypes to a dataset of 3871 previously-published ancient and present-day individuals. I used this dataset to carry out multiple quality control checks for the quality and preservation of the ancient data, and to assess the levels of contamination of each library with DNA from the present-day. The results of these checks were discussed with Kerttu Majander. Later, collaborators from the Max Planck Institute for Evolutionary Anthropology provided us with sequencing data from a present-day Saami individual that had already been processed. I incorporated these genotypes into the dataset and co-analysed them with the ancient data. I carried out all population genetic analyses (with the exception of the Y-chromosome haplotyping which was carried out by Choongwon Jeong) and discussed the results primarily with Kerttu Majander and Stephan Schiffels, and less frequently with a larger subset of coauthors.
I wrote the majority of the manuscript, with considerable contributions from Kerttu Majander and Stephan Schiffels, and input from all other co-authors. During the review process, development and implementation of new methodology, as well as testing of one method used during quality control was requested, which I carried out in full.

Initiation of the project	0%
Sample procurement	0%
Key Ideas for the project	50%
Laboratory work	0%
Bioinformatic data processing	90%
Method development, implementation and testing	100%
Population genetic analysis	95%
Manuscript writing	60%

3.2 Manuscript B

"Insights into South American population history from ancient genomes from Tierra del Fuego."

Thiseas C. Lamnidis, Zuzana Faltyskova, Christiana Lyn Scheib, Yali Xue, Hannes Schroeder, Susana Morano, Alfredo Prieto, Yolanda Espinosa-Parrilla, Carles Lalueza-Fox, Maru Mormina, Richard Durbin, Toomas Kivisild, Stephan Schiffels

Manuscript in preparation

In Manuscript B, we report whole genomes of 14 ancient individuals from Patagonia and Tierra del Fuego. One of these individuals has been previously published. We analyse these genomes together with 7 previously published ancient genomes from Tierra del Fuego, against a panel of present-day reference populations and relevant ancient individuals from South America. We identify a genetic shift taking place between 4,500 yBP and 1,100 yBP in Tierra del Fuego, followed by genetic continuity until at least 200 yBP. This shift entails the introduction of a genetic component previously identified in a 4,200 year-old population from Peru, which is

maximised in a 5,200 year-old population from San Nicolas island in California. We use rare variation to build a demographic model of modern Native American populations. We infer a split time between Siberians and Native Americans at 17,000 yBP, with a subsequent separation of Central and South American lineages at 10,000 yBP. Using rare allele sharing with the different modern (super-)populations in our demographic model, we infer the most likely branching point for each ancient individual. Inferred branching positions of ancient Fuego-Patagonian individuals are consistently close to the split time between the different Native American populations, implying a generalised affinity to present-day populations.

Author contributions thus far:

- These authors will have contributed equally to the future publication: Thiseas C. Lamnidis, Zuzana Faltyskova

Maru Mormina, Zuzana Faltyskova, Toomas Kivisild, and Richard Durbin designed the study. Yolanda Espinosa-Parrilla, Carles Lalueza-Fox, Alfredo Prieto, Susana Morano, Toomas Kivisild, and Richard Durbin collected the archaeological material. Zuzana Faltyskova, undertook all laboratory procedures, including sampling of the material, DNA extraction and library preparation. The libraries were sequenced at the University of Cambridge Department of Biochemistry DNA Sequencing Facility, and the Sanger Institute. I was provided with the sequencing data from 17 ancient Fuego-Patagonian individuals. I processed and quality controlled these data, together with sequencing data from 17 previously published individuals. The processing entailed mapping the raw sequencing data to the human reference genome, and verifying the ancient origin of the DNA by quantifying the characteristic aDNA deamination patterns present in the sequences. I ran multiple quality controls, by estimating both nuclear and mitochondrial DNA contamination in the data, and imposing strict coverage cut-offs. 14 of the ancient genomes in this study passed quality control checks, 5 of which had an average genomic coverage above 1x. I analysed the genomic data, and results were discussed primarily with Stephan Schiffels, and in less frequent intervals with a subset of the coauthors. I wrote the majority of the presented manuscript, with input from Stephan Schiffels, and minor comments from other coauthors.

Initiation of the project	0%
Sample procurement	0%
Key Ideas for the project	70%
Laboratory work	0%
Bioinformatic data processing	100%
Population genetic analysis	100%
Manuscript writing	75%

3.3 Manuscript C

"Origin and health status of first-generation Africans from early Colonial Mexico."

Rodrigo Barquera, Thiseas C. Lamnidis, Aditya Kumar Lankapalli, Arthur Kocher, Diana I. Hernández-Zaragoza, Elizabeth A. Nelson, Adriana C. Zamora-Herrera, Patxi Ramallo, Natalia Bernal-Felipe, Alexander Immel, Kirsten Bos, Víctor Acuña-Alonzo, Chiara Barbieri, Patrick Roberts, Alexander Herbig, Denise Kühnert, Lourdes Márquez-Morfín, Johannes Krause.

Accepted in Current Biology, 02 April 2020

In Manuscript C, we use a combined bioarchaeological approach, combining genetic data, osteological data, strontium isotope data from tooth enamel, δ13C and δ15N isotope data from dentine, and ethnohistorical information to reveal unprecedented detail on their origins and health of three putatively enslaved Africans from a mass burial at the *Hospital Real de San José de los Naturales* in Mexico City. Results from osteological analyses, radioisotope analyses, and genetic analyses of uniparental markers, genome-wide variation, and immunogenetic markers are all consistent with a Sub-Saharan African origin for the three analysed individuals. Osteological analyses of the three individuals reveal evidence suggesting a life experience of conflict and hardship. To identify possible pathogen infections that may have caused the death of these individuals, a screening for aDNA from various human pathogens was conducted. Complete genomes of *Treponema pallidum sub. pertenue* (causative agent of yaws) and Hepatitis B Virus (HBV) were recovered from these individuals, both of which were phyogenetically linked to strains from Sub-Saharan Africa. In concert, these results allow us to shed light into the lives and deaths of three first-generation enslaved Africans who died very early during the Colonial period in Mexico.

Author contributions:

- These authors contributed equally: Rodrigo Barquera, Thiseas C. Lamnidis, Aditya Kumar Lankapalli, Arthur Kocher, Diana I. Hernández-Zaragoza

In this project, while co-analysing the human nuclear DNA with Rodrigo Barquera, I also provided him with training and supervision in bioinformatics and population genetics. This project took a multi-disciplinary approach, including genetic, osteological, palaeopathological, and isotopic analyses. My involvement was limited to the processing, quality control and population genetic analysis of the human nuclear DNA.

Rodrigo Barquera undertook the sampling of the archaeological material, as well as DNA extraction and library preparation procedures for this project. The technical staff of the Max Planck Institute for the Science of Human History undertook the enrichment of these libraries for human DNA and the sequencing of the captured product. As part of the provided training, Rodrigo Barquera and I undertook quality control of the ancient human genetic data as well as population genetic analyses of the human nuclear DNA together. The results of these quality controls and analyses were extensively discussed between Rodrigo Barquera and me, and to a lesser extent with other co-authors involved in the population genetics analyses. I provided direction in the population genetic analyses, as well as insights into the limitations of different

methods. Additionally, I carried out Y-chromosome haplotyping on these individuals, the results of which were interpreted by Chiara Barbieri. Aditya Kumar Lankapalli carried out the analysis of treponemal aDNA that was identified in the genomic libraries from individual SJN003. Arthur Kocher analysed the viral aDNA identified in genomic libraries from individual SJN001. Osteological and palaeopathological analyses were performed by Rodrigo Barquera, Diana Hernández-Zaragoza, Elizabeth Nelson, Adriana Zamora-Herrera. Patrick Roberts and Patxi Ramallo performed the isotope analyses. All authors discussed the results. Rodrigo Barquera, Thiseas C. Lamnidis, Diana I. Hernández-Zaragoza, Elizabeth A. Nelson, Lourdes Márquez-Morfín and Johannes Krause wrote the paper with contributions from all authors. I contributed broadly to the manuscript and more specifically to sections pertaining to the population genetic analyses of the nuclear genomes and their interpretation. All authors read and approved the final version. Rodrigo Barquera, Diana I. Hernández-Zaragoza, Natalia Bernal-Felipe and Lourdes Márquez-Morfín wrote and submitted the final report (in Spanish) for the Council of Archaeology (INAH) based on the present manuscript.

Initiation of the project	0%
Sample procurement	0%
Key Ideas for the project	20%
Osteological/Isotope analyses	0%
Laboratory work	0%
Bioinformatic data processing	50%
Pathogenic agent analyses	0%
Population genetic analysis	70%
Manuscript writing	33%

4. Manuscript A

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-07483-5 OPEN

Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe

Thiseas C. Lamnidis[1], Kerttu Majander[1,2,3], Choongwon Jeong[1,4], Elina Salmela[1,3], Anna Wessman[5], Vyacheslav Moiseyev[6], Valery Khartanovich[6], Oleg Balanovsky[7,8,9], Matthias Ongyerth[10], Antje Weihmann[10], Antti Sajantila[11], Janet Kelso[10], Svante Pääbo[10], Päivi Onkamo[3,12], Wolfgang Haak[1], Johannes Krause[1] & Stephan Schiffels[1]

European population history has been shaped by migrations of people, and their subsequent admixture. Recently, ancient DNA has brought new insights into European migration events linked to the advent of agriculture, and possibly to the spread of Indo-European languages. However, little is known about the ancient population history of north-eastern Europe, in particular about populations speaking Uralic languages, such as Finns and Saami. Here we analyse ancient genomic data from 11 individuals from Finland and north-western Russia. We show that the genetic makeup of northern Europe was shaped by migrations from Siberia that began at least 3500 years ago. This Siberian ancestry was subsequently admixed into many modern populations in the region, particularly into populations speaking Uralic languages today. Additionally, we show that ancestors of modern Saami inhabited a larger territory during the Iron Age, which adds to the historical and linguistic information about the population history of Finland.

[1] Department of Archaeogenetics, Max Planck Institute for the Science of Human History, 07745 Jena, Germany. [2] Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, 72070 Tübingen, Germany. [3] Department of Biosciences, University of Helsinki, PL 56 (Viikinkaari 9), 00014 Helsinki, Finland. [4] The Eurasia3angle Project, Max Planck Institute for the Science of Human History, 07745 Jena, Germany. [5] Department of Cultures, Archaeology, University of Helsinki, PL 59 (Unioninkatu 38), 00014 Helsinki, Finland. [6] Peter the Great Museum of Anthropology and Ethnography (Kunstkamera), Russian Academy of Sciences, University Embankment, 3, Saint Petersburg 199034, Russia. [7] Vavilov Institute of General Genetics, Ulitsa Gubkina, 3, Moscow 117971, Russia. [8] Research Centre for Medical Genetics, Moskvorech'ye Ulitsa, 1, Moscow 115478, Russia. [9] Biobank of North Eurasia, Kotlyakovskaya Ulitsa, 3 строение 12, Moscow 115201, Russia. [10] Max Planck Institute for Evolutionary Anthropology, Deutscher Pl. 6, 04103 Leipzig, Germany. [11] Department of Forensic Medicine, University of Helsinki, PL 40 (Kytösuontie 11), Helsinki 00014, Finland. [12] Department of Biology, University of Turku, Turku 20014, Finland. These authors contributed equally to this work: Thiseas C. Lamnidis, Kerttu Majander. These authors jointly supervised this work: Päivi Onkamo, Wolfgang Haak, Johannes Krause, Stephan Schiffels. Correspondence and requests for materials should be addressed to P.O. (email: paivi.onkamo@helsinki.fi) or to W.H. (email: haak@shh.mpg.de) or to S.S. (email: schiffels@shh.mpg.de)

The genetic structure of Europeans today is the result of several layers of migration and subsequent admixture. The incoming source populations no longer exist in unadmixed form, but have been identified using ancient DNA in several studies over the last few years[1–8]. Broadly, present-day Europeans have ancestors in three deeply diverged source populations: European hunter-gatherers who settled the continent in the Upper Paleolithic, Europe's first farmers who expanded from Anatolia across Europe in the early Neolithic starting around 8000 years ago, and groups from the Pontic Steppe that arrived in Europe during the final Neolithic and early Bronze Age ~4500 years ago. As a consequence, most Europeans can be modelled as a mixture of these three ancestral populations[3].

This model, however, does not fit well for present-day populations from north-eastern Europe such as Saami, Russians, Mordovians, Chuvash, Estonians, Hungarians, and Finns: they carry additional ancestry seen as increased allele sharing with modern East Asian populations[1,3,9,10]. The origin and timing of this East Asian-related contribution is unknown. Modern Finns are known to possess a distinct genetic structure among today's European populations[9,11,12], and the country's geographical location at the crossroads of eastern and western influences introduces a unique opportunity to investigate the migratory past of north-east Europe. Furthermore, the early migrations and genetic origins of the Saami people in relation to the Finnish population call for a closer investigation. Here, the early-Metal-Age, Iron-Age, and historical burials analysed provide a suitable time-transect to ascertain the timing of the arrival of the deeply rooted Siberian genetic ancestry, and a frame of reference for investigating linguistic diversity in the region today.

The population history of Finland is subject to an ongoing discussion, especially concerning the status of the Saami as the earlier inhabitants of Finland, compared to Finns. The archaeological record proves human presence in Finland since 9000 BC[13]. Over the millennia, people from Scandinavia, the north-east Baltics, and modern-day Russia have left evidence of their material cultures in Finland[14,15]. The Finno-Ugric branch of the Uralic language family, to which both Saami and Finnish languages belong, has diverged from other Uralic languages no earlier than 4000–5000 years ago, when Finland was already inhabited by speakers of a language today unknown. Linguistic evidence shows that Saami languages were spoken in Finland prior to the arrival of the early Finnish language and have dominated the whole of the Finnish region before 1000 CE[16–18]. Particularly, southern Ostrobothnia, where Levänluhta is located, has been suggested through place names to harbour a southern Saami dialect until the late first millennium[19], when early Finnish took over as the dominant language[20]. Historical sources note Lapps living in the parishes of central Finland still in the 1500s[21]. It is, however, unclear whether all of them spoke Saami, or if some of them were Finns who had changed their subsistence strategy from agriculture to hunting and fishing. There are also documents of intermarriage, although many of the indigenous people retreated to the north (see ref. [22] and references therein). Ancestors of present-day Finnish speakers possibly migrated from northern Estonia, to which Finns still remain linguistically close, and displaced but also admixed with the local population of Finland, the likely ancestors of today's Saami speakers[23].

In this study, we present new genome-wide data from Finland and the Russian Kola Peninsula, from 11 individuals who lived between 3500 and 200 years ago (and 4 more ancient genomes with very low coverage). In addition, we present a new high-coverage whole genome from a modern Saami individual for whom genotyping data was previously published[1]. Our results suggest that a new genetic component with strong Siberian affinity first arrived in Europe at least 3500 years ago, as observed in our oldest analysed individuals from northern Russia. These results describe the gene pool of modern north-eastern Europeans in general, and of the speakers of Uralic languages in particular, as the result of multiple admixture events between Eastern and Western sources since the first appearance of this ancestry component. Additionally, we gain further insights into the genetic history of the Saami in Finland, by showing that during the Iron Age, close genetic relatives of modern Saami lived in an area much further south than their current geographic range.

Results

Sample information and archaeological background. The ancient individuals analysed in this study come from three time periods (Table 1, Fig. 1, Supplementary Note 1). The six early Metal Age individuals were obtained from an archaeological site at Bolshoy Oleni Ostrov in the Murmansk Region on the Kola Peninsula (Bolshoy from here on). The site has been radiocarbon dated to 1610-1436 calBCE (see Supplementary Note 1) and the mitochondrial DNA HVR-I haplotypes from these six individuals have been previously reported[24]. Today, the region is inhabited by Saami. Seven individuals stem from excavations in Levänluhta, a lake burial in Isokyrö, Finland. Artefacts from the site have been dated to the Finnish Iron Age (300–800 CE)[25,26]. Today, the inhabitants of the area speak Finnish and Swedish. Two individuals were obtained from the 18–19th century Saami cemetery of Chalmny Varre on the Russian Kola Peninsula. The cemetery and the surrounding area were abandoned in the 1960s because of planned industrial constructions, and later became the subject of archaeological excavations. In addition, we sequenced the whole genome of a modern Saami individual to 17.5-fold coverage, for whom genotyping data has previously been published[1].

The sampling and subsequent processing of the ancient human remains was done in dedicated clean-room facilities (Methods). A SNP-capture approach targeting a set of 1,237,207 single nucleotide polymorphisms (SNPs) was used to enrich ancient-DNA libraries for human DNA[4]. The sequenced DNA fragments were mapped to the human reference genome, and pseudohaploid genotypes were called based on a random read covering each targeted SNP (see Methods). To ensure the ancient origin of our samples, and the reliability of the data produced, we implemented multiple quality controls. First, we confirmed the deamination patterns at the terminal bases of DNA reads being characteristic of ancient DNA (Supplementary Table 1). Second, we estimated potential contamination rates through heterozygosity on the single-copy X chromosome for male individuals (all below 1.6%)[27] (Table 1, see Supplementary Figure 1 for sex determination). Third, we carried out supervised genetic clustering using ADMIXTURE[28], using six divergent populations as defined clusters, to identify genetic dissimilarities and possible contamination from distantly related sources (Supplementary Figure 2, Supplementary Note 2). Fourth, we calculated mitochondrial contamination for all ancient individuals using ContamMix[29]. Finally, for individuals with sufficient SNP coverage, we carried out principal component analysis (PCA), projecting damage-filtered and non-filtered versions of each individual, to show that the different datasets cluster together regardless of the damage-filtering[30] (see Supplementary Figure 3 and Methods). Eleven ancient individuals passed those quality checks, while four individuals from Levänluhta were excluded from further analyses, due to low SNP coverage (<15,000 SNPs). We merged the data from these 15 individuals with that of 3333 published present-day individuals genotyped on the Affymetrix Human Origins platform and 538 ancient individuals sequenced using a mixture of DNA capture and shotgun sequencing (Supplementary Data 1)[1–4,8,31,32].

Table 1 Sample information

Individual ID	Site/ location	Date	Population label	Genetic Sex	# SNPs overlap with Human Origins	Avg. coverage on target (1240 K MapQ ≥ 30)	Nuclear contamination	mtDNA contamination	mtDNA, Y haplotypes
BOO001	Bolshoy Oleni Ostrov, Murmansk, Russia	3473 ± 87 calBP	Bolshoy	F	347,709	1.22	N/A	0.001 (0.000–0.007)	U4a1
BOO002			Bolshoy	M	403,994	0.81	0.004 ± 0.003	0.000 (0.000–0.002)	Z1a1a, N1c1a1a
BOO003			Bolshoy	F	300,598	1.09	N/A	0.016 (0.010–0.024)	T2d1b1
BOO004			Bolshoy	M	342,582	2.92	0.002 ± 0.001	0.003 (0.000–0.059)	C4b, N1c1a1a
BOO005			Bolshoy	F	347,042	2.88	N/A	0.005 (0.001–0.026)	U5a1d
BOO006			Bolshoy	F	390,835	1.08	N/A	0.008 (0.004–0.014)	D4e4
CHV001	Chalmny- Varre, Murmansk, Russia	18–19th cent CE	Chalmny Varre	F	426,702	1.42	N/A	0.001 (0.000–0.003)	U5b1b1a3
CHV002			Chalmny Varre	M	215,228	0.47	0.016 ± 0.011	0.000 (0.000–0.003)	V7a1, I2a1
JK1963	Levänluhta, Isokyrö, Finland	300–800 CE	N/A	f	10,492	0.02	N/A	0.004 (0.001–0.013)	N/A
JK1967			N/A	f	5267	0.01	N/A	0.000 (0.000–0.009)	N/A
JK1968			Levänluhta	F	207,076	0.47	N/A	0.004 (0.001–0.065)	U5a1a1a'b'n
JK1970			Levänluhta	F	194,764	0.44	N/A	0.023 (0.005–0.129)	U5a1a1
JK2065			Levänluhta_B	F	133,224	0.29	N/A	0.028 (0.012–0.064)	K1a4a1b
JK2066			N/A	?	4045	0.01	N/A	0.002 (0.001–0.007)	N/A
JK2067			N/A	?	2356	0.01	N/A	0.010 (0.002–0.032)	N/A
Saami001	Finland	Modern	Modern Saami	M	593,094	6.32	N/A	N/A	U5b1b1a1, I1a1b3a1

Summary information for all individuals for which we report genomic data in this study. A radiocarbon date is given for the Bolshoy samples, as described in Supplementary Text 1. Other dates are context-based, as described in Supplementary Text 1. Population labels for individuals of low coverage are shown as N/A. Genetic sex is determined as described in Methods, and lowercase letters denote probable genetic sex for low-coverage individuals. Question marks denote undefined genetic sex due to low coverage. Mitochondrial contamination estimates are given as posterior mode and 95% posterior intervals in brackets, as provided by ContamMix[70]

Eastern genetic affinities in Northern Europe. To investigate the genetic affinities of the sampled individuals, we projected them onto principal components (PC) computed from 1320 modern European and Asian individuals (Fig. 2a, Supplementary Figures 3a, b for a version focusing on West Eurasia). As expected, PC1 separates East Asian from West Eurasian populations. Within each continental group, genetic variability is spread across PC2: The East Asian genetic cline contains populations between the Siberian Nganasan (Uralic speakers) and Yukagirs at one end, and the Ami and Atayal from Taiwan at the other end. The West Eurasian cline along PC2 spans from the Bedouins on the Arabian Peninsula to north-eastern Europeans including Lithuanians, Norwegians and Finns. Between these two main Eurasian clines exist multiple clines, spanning between West and East Eurasians. These clines are likely the result of admixture events and population movements between East and West Eurasia. Most relevant to the populations analysed here is the admixture cline between north-eastern Europe and the North Siberian Nganasan, including mostly Uralic-speaking populations in our dataset (marked in light purple in Fig. 2a).

Ten of the eleven ancient individuals from this study fall on this Uralic cline, with the exception of one individual from Levänluhta (ID JK2065, here named Levänluhta_B), who instead is projected closer to modern Lithuanian, Norwegian and Icelandic populations. Specifically, two Levänluhta individuals and the two historical Saami from Russia are projected very close to the two previously published modern Saami (Saami.DG)[32] and the new Saami shotgun genome generated in this study (as well as the previously published genome of the same individual, here labelled Saami (WGA)[1]), suggesting genetic continuity in the north from the Iron Age to modern-day Saami populations. In contrast, the six ancient individuals from Bolshoy are projected much further towards East Asian populations, and fall to an intermediate position along the Uralic cline and close to modern-day Mansi.

Unsupervised genetic clustering analysis as implemented in the ADMIXTURE[28] program suggests a similar profile to the PCA: north-eastern European populations harbour a Siberian genetic component (light purple) maximized in the Nganasan (Fig. 2b, see Supplementary Figure 4a for results over multiple K values). The hunter-gatherer genetic ancestry in Europeans (blue) is maximized in European Upper Palaeolithic and Mesolithic hunter-gatherers, including the 8000-year-old Western European hunter-gatherers from Hungary and Spain (WHG), the 8000-year-old Scandinavian hunter-gatherers from Motala (SHG) and the Narva and Kunda individuals from the Baltics. An ancestry component associated with Europe's first farmers (orange) is maximized in Early Neolithic Europeans associated with the LBK (from German: Linearbandkeramik). The steppe ancestry component within modern Europeans (green), which is associated with the Yamnaya population, is maximised in ancient Iranian populations and to a lesser extent Caucasus hunter-gatherers (CHG). This ancestry component is also present in modern Armenians from the Caucasus, Bedouins from the Arabian Peninsula and South Asian populations. Within modern Europeans, the Siberian genetic component (light purple) is maximized in the Mari and Saami, and can also be seen in similar proportions in the historical Saami from Chalmny Varre and in two of the Levänluhta individuals. The third Levänluhta individual (Levänluhta_B), however, lacks this Siberian

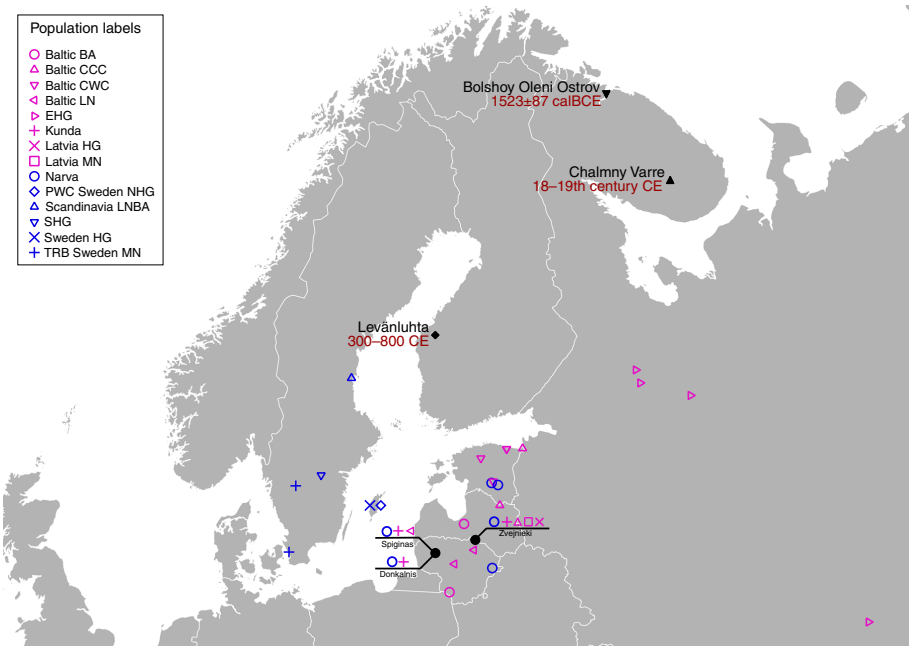


Fig. 1 Location and age of archaeological sites used in this study. The location of other sites relevant to this study is also shown. Markers used correspond to the ones used in Fig. 2. Map generated with QGIS 2.18.19 (http://www.qgis.org/) using the Natural Earth country boundary dataset (http://www.naturalearthdata.com) for the basemap. Source data are provided as a Source Data file

component. The six ancient individuals from Bolshoy show substantially higher proportions of the Siberian component: it comprises about half of their ancestry (42.3–58.2%), whereas the older Mesolithic individuals from Motala (SHG) do not possess it at all. The Native-American-related ancestry seen in the EHG and Bolshoy corresponds to a previously reported affinity towards Ancient North Eurasians (ANE)[2,33] contributing genes to both Native Americans and West Eurasians. ANE ancestry also comprises part of the ancestry of Nganasans[2].

Interestingly, results from uniparentally-inherited markers (mtDNA and Y chromosome) as well as certain phenotypic SNPs also show Siberian signals in Bolshoy: mtDNA haplogroups Z1, C4 and D4, common in modern Siberia[24,34,35] are represented by the individuals BOO002, BOO004 and BOO006, respectively (confirming previous findings[24]), whereas the Y-chromosomal haplotype N1c1a1a (N-L392) is represented by the individuals BOO002 and BOO004. Haplogroup N1c, to which this haplotype belongs, is the major Y-chromosomal lineage in modern north-east Europe and European Russia. It is especially prevalent in Uralic speakers, comprising for example as much as 54% of eastern Finnish male lineages today[36]. Notably, this is the earliest known occurrence of Y-haplogroup N1c in Fennoscandia. Additionally, within the Bolshoy population, we observe the derived allele of rs3827760 in the *EDAR* gene, which is found in near-fixation in East Asian and Native American populations today, but is extremely rare elsewhere[37], and has been linked to phenotypes related to tooth shape[38] and hair morphology[39] (Supplementary Data 2). Scandinavian hunter-gatherers from Motala in Sweden have also been found to carry haplotypes associated with this allele[4]. Finally, in the Bolshoy individuals we also see high frequencies of haplotypes associated with diets rich in poly-unsaturated fatty acids, in the *FADS* genes[4,40,41].

The arrival of Siberian ancestry in Europe. We formally tested for admixture in north-eastern Europe by calculating f_3(Test; Siberian source, European source) statistics. We used several Uralic-speaking populations—Estonians, Saami, Finnish, Mordovians and Hungarians—and Russians as Test populations. Significantly negative f_3 values correspond to the Test population being admixed between populations related to the two source populations[42]. We used multiple European and Siberian sources to capture differences in ancestral composition among proxy populations: As proxies for the Siberian source we used Bolshoy, Mansi and Nganasan, and for the European source we used modern Icelandic, Norwegian, Lithuanian and French. Our results show that all of the test populations are indeed admixed, with the most negative values arising when Nganasan are used as the Siberian source (Supplementary Data 3). Among the European sources, Lithuanians gave the most negative results for Estonians, Russians and Mordovians. For modern Hungarians, the European source giving most negative results was French, while both Bolshoy and Nganasan gave equally negative results when used as the Siberian source. With Finns as test, modern Icelanders were the European source giving most negative statistics. Finally, Icelanders and Nganasan used as the European and Siberian sources, respectively, yielded the most negative result for the present-day Saami as a Test. This result is still non-significantly negative, either due to the low number of modern Saami individuals in our dataset ($n = 3$), or due to post-admixture drift in modern Saami. A high degree of population-specific drift can affect f_3-statistics and result in less negative and even positive values[42]. Indeed, post-admixture drift would correlate well with the suggested founder effect[43] in Saami. To further test differential relatedness with Nganasan in European populations and in the ancient individuals in this study, we

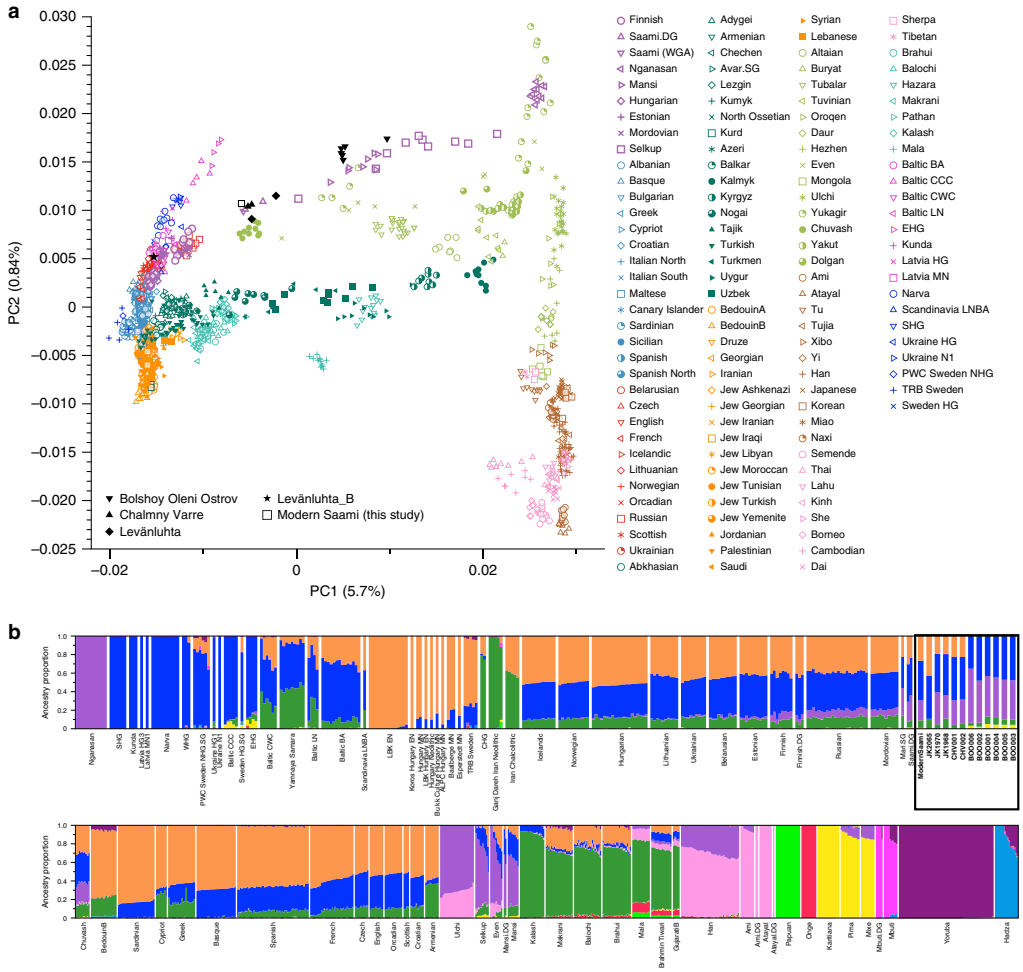


Fig. 2 PCA and ADMIXTURE analysis. **a** PCA plot of 113 Modern Eurasian populations, with individuals from this study and other relevant ancient genomes projected on the principal components, using the "shrinkmode: YES" option. Uralic-speaking populations are highlighted in dark purple. PCA of Europe can be found in Supplementary Figure 3. **b** Plot of ADMIXTURE (K = 11) results containing worldwide populations. Ancient individuals from this study (black box) are represented by thicker bars and shown in bold. For a figure of ADMIXTURE results over multiple K values see Supplementary Figure 4a. Source data are provided as a Source Data file

calculated f_4(Mbuti, Nganasan; Lithuanian, Test) (Fig. 3). Consistent with f_3-statistics above, all the ancient individuals and modern Finns, Saami, Mordovians and Russians show excess allele sharing with Nganasan when used as Test populations. Of all Uralic-speaking populations in Europe, Hungarians are the only population that shows no evidence of excess allele sharing with Nganasan compared to that of Lithuanians, consistent with their distinct population history from other Uralic populations as evidenced by historical sources (see ref. [44] and references therein).

We further estimated the genetic composition in these populations using qpAdm[3]. All ancient and modern individuals from the Baltics, Finland and Russia were successfully modelled as a mixture of five lines of ancestry, represented by eastern Mesolithic hunter-gatherers (EHG, from Karelia), Yamnaya from Samara, LBK from the early European Neolithic, western Mesolithic hunter-gatherers (WHG, from Spain, Luxembourg and Hungary), and Nganasan, or subsets of those five (Supplementary Data 4). In contrast to previous models for European populations using three streams of ancestry[2,3], we found that some populations modelled here require two additional components: a component related to modern Nganasans, as discussed above, and additional EHG ancestry, not explained by Yamnaya (who have been shown to contain large amounts of EHG ancestry themselves[3]). Indeed, the six Bolshoy individuals have substantial amounts of EHG but no Yamnaya ancestry. We find that Nganasan-related ancestry is significantly present in all of our ancient samples except for Levänluhta_B, and in many modern, mainly Uralic-speaking populations. The 3500-year-old ancient individuals from Bolshoy represent the highest proportion of Siberian Nganasan-related ancestry seen in this

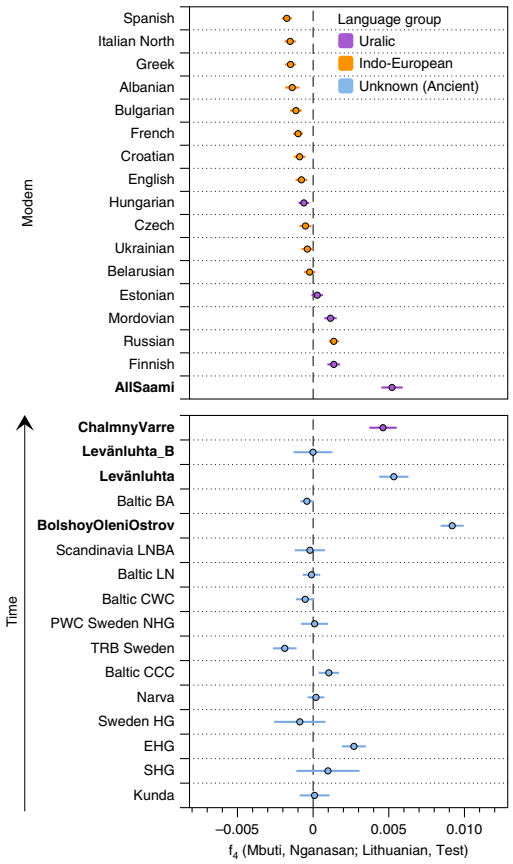


Fig. 3 Calculated f_4 (Mbuti, Nganasan; Lithuanian, Test). Modern populations are sorted by f_4 magnitude; ancient populations are sorted through time. Populations are coloured based on language family, with Ancient individuals with unknown language shown in blue. "AllSaami" refers to a grouping including the 2 individuals from the SGDP (Saami (SGDP)) and the high-coverage modern Saami shotgun genome in this study (Modern Saami). Individuals from this study are indicated by labels in bold. Error bars represent 3 standard errors, to indicate significant difference from 0. Source data are provided as a Source Data file

region so far, and possibly evidence its earliest presence in the western end of the trans-Siberian expanse (Fig. 4). The geographically proximate ancient hunter-gatherers from the Baltics (6000 and 6300 BC) and Motala (~6000 BC), who predate Bolshoy, lack this component, as do late Neolithic and Bronze Age individuals from the Baltics[7,8,45]. All later ancient individuals in this study have lower amounts of Nganasan-related ancestry than Bolshoy (Figs 3, 4), probably as a result of dilution through admixture with other populations from further south. This is also consistent with the increased proportion of early European farmer ancestry related to Neolithic Europeans (Fig. 2b) in our later samples. We note that a low but significant amount of Neolithic European ancestry is also present in the Bolshoy population. Finally, we tested whether Bolshoy, instead of Nganasan, can be used as source population. We found that Bolshoy works as source in some ancient individuals, but not for modern Uralic speakers (see Supplementary Data 4).

As shown by these multiple lines of evidence, the pattern of genetic ancestry observed in north-eastern Europe is the result of admixture between populations from Siberia and populations from Europe. To obtain a relative date of this admixture, and as an independent line of evidence thereof, we used admixture linkage disequilibrium decay, as implemented in ALDER[46]. Our ALDER admixture estimate for Bolshoy, using Nganasan and EHG as admixture sources, dates only 17 generations ago. Based on the radiocarbon date for Bolshoy and its uncertainty, and assuming a generation time of 29 years[47], we estimate the time of introduction of the Siberian Nganasan-related ancestry in Bolshoy to be 3977 (± 77) years before present (yBP) (Fig. 5b). Estimates obtained using Nganasans and Lithuanians as source populations provided a similar estimate (Supplementary Figure 5 for LD decay plots for multiple populations using Lithuanian and Nganasan as sources.). ALDER provides a relative date estimate for a single-pulse admixture event in generations. When multiple admixture events have occurred, such a single estimate should be interpreted as a (non-arithmetic) average of those events[46,48]. Therefore the admixture date estimate for Bolshoy does not preclude earlier admixture events bringing Siberian Nganasan-related ancestry into the population, in multiple waves. Indeed, for all other populations with evidence of this ancestry, we find much younger admixture dates (Fig. 5a), suggesting that the observed genetic ancestry in north-eastern Europe is inconsistent with a single-pulse admixture event.

Major genetic shift in Finland since the Iron Age. Besides the early evidence of Siberian ancestry, our ancient samples from Levänluhta and Chalmny Varre allow us to investigate the more recent population history of Finland. To test whether the ancient individuals from Levänluhta form a clade with modern-day Saami or with modern Finns, we calculated f_4(Saami(SGDP), Test; *X*, Mbuti) and f_4(Finnish, Test; *X*, Mbuti), where Test was substituted with each ancient individual from Levänluhta, the two historical Saami individuals from Chalmny Varre, as well as the Modern Saami individual, and *X* was substituted by worldwide modern-day populations (Supplementary Data 5 & 6, and Supplementary Figures 5 & 6). One Levänluhta (JK1968) and the two Chalmny Varre individuals consistently formed a clade with modern-day Saami, but not with modern-day Finns, with respect to all worldwide populations. One Levänluhta individual (JK1970) showed slightly lower affinity to central Europe than modern-day Saami do, while still rejecting a cladal position with modern-day Finns. This indicates that the people inhabiting Levänluhta during the Iron Age, and possibly other areas in the region as well, were more closely related to modern-day Saami than to present-day Finns; however, their difference from the modern Saami may reflect internal structure within the Saami population or additional admixture into the modern population.

One of the individuals from Levänluhta (JK2065/Levänluhta_B) rejects a cladal position with modern Saami to the exclusion of most modern Eurasian populations. This individual also rejects a cladal position with Finns. We analysed low coverage genomes from four additional individuals of the Levänluhta site using PCA (Supplementary Figure 3), confirming the exclusive position of Levänluhta_B compared to all other six individuals (including the four low-coverage individuals) from that site, as is consistent with the ADMIXTURE and qpAdm results. The outlier position of this individual cannot be explained by modern contamination, since it passed several tests for authentication (see Methods) along with all other ancient individuals. However, no direct dating was available for the Levänluhta material, and we cannot exclude the

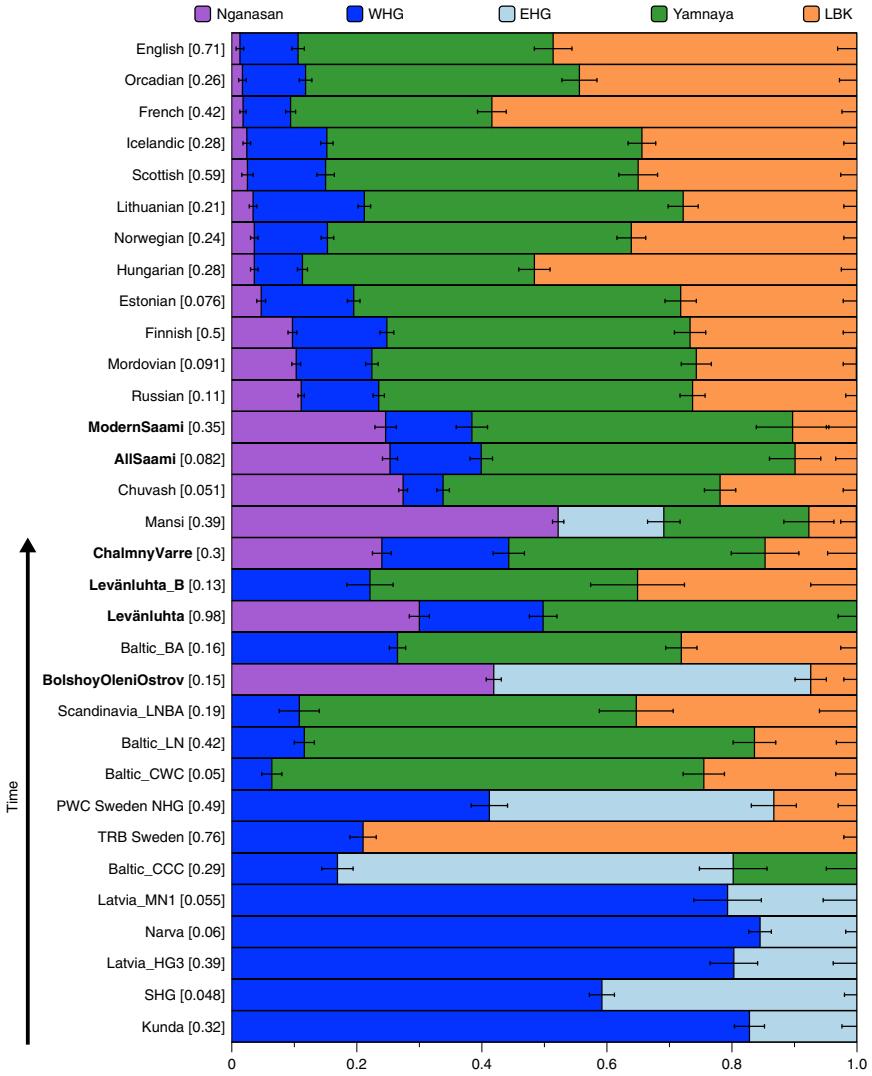


Fig. 4 Mixture proportions from five sources estimated using qpAdm. Sources used were Nganasan, WHG, EHG, Yamnaya and LBK (see Methods/Supplementary Data 4). *P*-values (chi-square) for each model are shown in square brackets next to the test population. Results from the least complex model for each test population/individual are shown. "AllSaami" corresponds to a population consisting of the two genomes from the SGDP and the genome from this study (ModernSaami). Error bars represent one standard error and are plotted to the right of their associated mixture proportion. Populations containing individuals from this study are shown in bold. Source data are provided as a Source Data file

possibility of a temporal gap between this individual and the other individuals from that site.

Discussion

In terms of ancient human DNA, north-eastern Europe has been relatively understudied. In this study we extend the available information from this area considerably, and present the first ancient genome-wide data from Finland. While the Siberian genetic component presented here has been previously described in modern-day populations from the region[1,3,9,10], we gain further insights into its temporal depth. Our data suggest that this fourth genetic component found in modern-day north-eastern Europeans arrived in the area before 3500 yBP. It was introduced in the population ancestral to Bolshoy Oleni Ostrov individuals 4000 years ago at latest, as illustrated by ALDER dating using the ancient genome-wide data from the Bolshoy samples. The upper bound for the introduction of this component is harder to estimate. The component is absent in the Karelian hunter-gatherers (EHG)[3] dated to 8300–7200 yBP as well as Mesolithic and

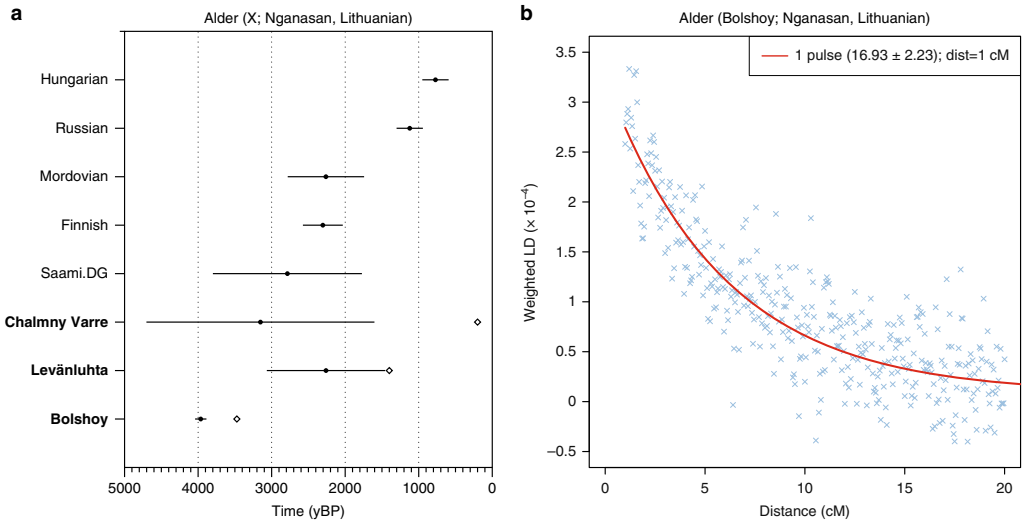


Fig. 5 Dating the introduction of Siberian ancestry using ALDER. **a** ALDER-inferred admixture dates (filled circles) for different populations, using Nganasan and Lithuanian as sources. Available dates for ancient populations are shown in white diamonds. Error bars represent one standard error provided by ALDER and include the uncertainty surrounding the dating of ancient population samples, calculated using standard propagation. Populations containing individuals from this study are shown in bold. **b** LD decay curve for Bolshoy, using Nganasan and EHG as sources. The fitted trendline considers a minimum distance of 1 cM. A full set of LD decay plots can be found in Supplementary Figure 5. Source data are provided as a Source Data file

Neolithic populations from the Baltics from 8300 yBP and 7100–5000 yBP respectively[8]. While this suggests an upper bound of 5,000 yBP for the arrival of this Siberian ancestry, we cannot exclude the possibility of its presence even earlier, yet restricted to more northern regions, as suggested by its absence in populations in the Baltics during the Bronze Age. Furthermore, our study presents the earliest occurrence of the Y-chromosomal haplogroup N1c in Fennoscandia. N1c is common among modern Uralic speakers, and has also been detected in Hungarian individuals dating to the 10th century[44], yet it is absent in all published Mesolithic genomes from Karelia and the Baltics[3,8,45,49].

The large Nganasan-related component in the Bolshoy individuals from the Kola Peninsula provides the earliest direct genetic evidence for an eastern migration into this region. Such contact is well documented in archaeology, with the introduction of asbestos-mixed Lovozero ceramics during the second millennium BC[50], and the spread of even-based arrowheads in Lapland from 1900 BCE[51,52]. Additionally, the nearest counterparts of Vardøy ceramics, appearing in the area around 1,600-1,300 BCE, can be found on the Taymyr peninsula, much further to the East[51,52]. Finally, the Imiyakhtakhskaya culture from Yakutia spread to the Kola Peninsula during the same period[24,53]. Contacts between Siberia and Europe are also recognised in linguistics. The fact that the Nganasan-related genetic component is consistently shared among Uralic-speaking populations, with the exceptions of absence in Hungarians and presence in the non-Uralic speaking Russians, makes it tempting to equate this genetic component with the spread of Uralic languages in the area. However, such a model may be overly simplistic. First, the presence of the Siberian component on the Kola Peninsula at ca. 3500 yBP predates most linguistic estimates of the spread of extant Uralic languages to the area[54]. Second, as shown in our analyses, the admixture patterns found in historic and modern Uralic speakers are complex and in fact inconsistent with a single admixture event. Therefore, even if the Siberian genetic component partly spread alongside Uralic languages, some Siberian ancestry may have been already present in the area from earlier admixture events.

The novel genome-wide data presented here from ancient individuals from Finland opens new insights into Finnish population history. Two of the three higher-coverage genomes and all four low-coverage genomes from Levänluhta individuals showed low genetic affinity to modern-day Finnish speakers of the area. Instead, an increased affinity was observed to modern-day Saami speakers, now mostly residing in the north of the Scandinavian Peninsula. These results suggest that the geographic range of the Saami extended further south in the past, and points to a genetic shift at least in the western Finnish region since the Iron Age. The findings are in concordance with the noted linguistic shift from Saami languages to early Finnish. Further ancient DNA from Finland is needed in order to conclude, to what extent these signals of migration and admixture are representative of Finland as a whole.

Methods

Sampling. Written informed consent was obtained from the Saami individual whose genome was analysed in this study, which was approved of by The Hospital District of Helsinki and Uusimaa Ethical Committee (Decision 329/13/03/00/2013) and the ethics committee of the University of Leipzig (Approval reference number 398-13-16122013).

Sampling and extracting ancient DNA requires a strict procedure in order to avoid contamination introduced by contemporary genetic material. For the 13 Iron-Age individuals from Finland available to us, the sampling took place in a clean-room facility dedicated to ancient DNA work, at the Institute for Archaeological Sciences in Tübingen. The preliminary workflow included documenting, photographing and storing the samples in individual, ID-coded plastic tubes and plastic bags. As a result of an early pilot study, the tooth samples we used were fragmented, and some of the dentine was removed. The remaining dentine was collected by carefully separating it from the enamel with a dentist drill and cooled-down diamond drill heads, rotated at a speed below 15 rpm, to avoid possible heat-caused damaging to the ancient DNA.

For samples from the sites of Bolshoy and Chalmny Varre, we used leftover tooth powder that was originally processed at the Institute of Anthropology at the

University of Mainz for replication purposes as described in ref. [24]. In brief, the sample preparation steps included UV-irradiation for 30–45 min, followed by gentle wiping of the surface with diluted commercial bleaches. The teeth were then sand-blasted using aluminium oxide abrasive (Harnisch & Rieth) and ground to fine powder using a mixer-mill (Retsch).

Radiocarbon date calibration. We calibrated the radiocarbon date of Bolshoy, reported in refs [24,55] as 3473 ± 42 years BP, using Intcal13[56] as the calibration curve, using OxCal 4.3[57].

DNA extraction and library preparation. DNA from the six Bolshoy and the two Chalmny Varre samples was extracted in the ancient DNA facilities of the Max Planck Institute for the Science of Human History (MPI-SHH) in Jena, Germany. Extraction for the Levänluhta samples was similarly conducted in the clean-room facilities of the Institute for Archaeological Sciences in Tübingen. For each specimen, ~ 50 mg of dentine powder was used for an extraction procedure specifically designed for ancient DNA retrieval[58]. Extraction buffer containing 0.45 M EDTA, pH 8.0 (Life Technologies) and 0.25 mg/ml Proteinase K (Sigma-Aldrich) was added to bone powder and incubated at 37 °C with rotation overnight. The supernatant was separated from the pellet of bone powder by centrifugation (14,000 rpm). A binding buffer consisting of 5 M GuHCL (Sigma Aldrich) and 40% Isopropanol (Merck), together with 400 µl of 1 M sodium acetate (pH 5.5) was added to the supernatant, and the solution purified by spinning it through a purification column attached to a High Pure Extender Assembly funnel (8 min. in 1500 rpm, with slow acceleration). The column was then spun into a collection tube (1 min 14,000 rpm) 1–2 times to maximise the yield. This was followed by two subsequent washing steps of 450 µl of wash buffer (High Pure Viral Nucleic Acid Large Volume Kit) and two dry spin steps of 1 min centrifugation at 14,000 rpm. The final total volume of 100 µl eluate was reached by two separate elution rounds of 50 µl of TET (10 mM Tris-HCL, 1 mM EDTA pH 8.0, 0.1% Tween20), each spun for 1 min at 14,000 rpm into a fresh Eppendorf 1.5 ml tube. Negative controls (buffer instead of sample) were processed in parallel at a ratio of 1 control per 7 samples.

Of the 100 µl extract, 20 µl was used to immortalize the sample DNA as a double-stranded library. The procedure included a blunt-end repair, adapter ligation and adapter fill-in steps, as described by Meyer and Kircher[59]. During the blunt-end repair step, a mixture of 0.4 U/µl T4 PNK (polynucleotide kinase) and 0.024 U/µl T4 DNA polymerase, 1× NEB buffer 2 (NEB), 100 µM dNTP mix (Thermo Scientific), 1 mM ATP (NEB) and 0.8 mg/ml BSA (NEB) was added to the template DNA, followed by incubation in a thermocycler (15 min 15 °C, 15 min 25 °C) and purification with a MinElute kit (QIAGEN). The product was eluted in 18 µl TET buffer. The adapter ligation step included a mixture of 1× Quick Ligase Buffer (NEB), 250 nM Illumina Adapters (Sigma-Aldrich) and 0.125 U/µl Quick Ligase (NEB), added to the 18 µl eluate, followed by a 20 min incubation, and second purification step with MinElute columns, this time in 20 µl eluate. For the fill-in step, a mixture of 0.4 U/µl Bst-polymerase and 125 µM dNTP mix was added and the mixture then incubated in a thermocycler (30 min 37 °C, 10 min 80 °C). Libraries without Uracil-DNA-glycosylase (UDG) treatment were produced for all of the 13 extracts from Levänluhta. In addition, UDG-half treated libraries were produced for seven of the original 13 extracts from Levänluhta, and for all Bolshoy and Chalmny Varre extracts. To introduce the UDG-half treatment, an initial stage was included in the library preparation, in which 250 U USER enzyme (NEB) was added into the 20 µl of extract, followed by an incubation at 37 °C for 30 min, and then 12 °C for 1 min. This again was followed by the addition of 200 U UGI (Uracil Glycosylase inhibitor, by NEB) and another identical incubation to stop the enzymatic excision of deaminated sites, as described in[60]. For each library, a unique pair of eight-bp-long indexes was incorporated using a Pfu Turbo Cx Hotstart DNA Polymerase and a thermocycling program with the temperature profile as follows: initial denaturation (98 °C for 30 sec), cycle of denaturation/annealing/elongation (98 °C for 10 sec/ 60 °C for 20 sec/ 72 °C for 20 sec) and final extension at 72 °C for 10 min[61]. Bone powder from a cave bear was processed in parallel serving as a positive control. Negative controls for both extraction and library preparation stages were kept alongside the samples throughout the entire workflow.

Experiment efficiency was ensured by quantifying the concentration of the libraries on qPCR (Roche) using aliquots from libraries before and after indexing. The molecular copy number in pre-indexed libraries ranged from ~ 10E8 to ~ 10E9 copies/µl, indicating a successful library preparation, whereas the indexed libraries ranged from ~ 10E10 to ~ 10E12 copies/µl, stating an admissible indexing efficiency. The negative controls showed 4–5 orders of magnitude lower concentration than the samples, indicating low contamination levels from the laboratory processing stages.

The libraries were amplified with PCR, for the amount of cycles corresponding to the concentrations of the indexed libraries, using AccuPrime Pfx polymerase (5 µl of library template, 2 U AccuPrime Pfx DNA polymerase by Invitrogen, 1 U of readymade 10× PCR mastermix, and 0.3 µM of primers IS5 and IS6, for each 100 µl reaction) with thermal profile of 2 min denaturation at 95 °C, 3–9 cycles consisting of 15 sec denaturation at 95 °C, 30 sec annealing at 60 °C, 2 min elongation at 68 °C and 5 min elongation at 68 °C. The amplified libraries were purified using MinElute spin columns with the standard protocol provided by the manufacturer (Qiagen), and quantified for sequencing using an Agilent 2100 Bioanalyzer DNA 1000 chip.

For the modern Saami individual, total DNA was phenol-chloroform extracted and physically sheared using COVARIS fragmentation. A modified Illumina library preparation was performed using blunt-end repair followed by A-tailing of the 3'-end and ligation of forked adapters. Indexing PCR was followed by excision of fragments ranging from 500 to 600 bp from a 2% agarose gel.

Capture & sequencing. We used the in-solution capture procedure from ref. [62] to enrich our libraries for DNA fragments overlapping with 1,237,207 variable positions in the human genome[4]. The sequences used as bait, attached to magnetic beads, were mixed in with the DNA sample template in solution, and left to hybridize with the target DNA during a 24-hour incubation at 60 °C in a rotation oven. 4–6 samples were pooled in equal mass ratios at a total of 2000 ng of DNA. The captured libraries were sequenced (75 bp single-end, plus additional paired-end for the three non-UDG libraries of the Levänluhta individuals) on an Illumina HiSeq 4000 platform at the Max Planck Institute for the Science of Human History in Jena. Out of the 13 originally processed Iron-Age samples from Finland, seven proved to be of an adequate quality to be used in downstream analyses. The modern Saami genome was sequenced in on a Genome Analyser II (8 lanes, 125 bp paired-end) at the Max Planck Institute for Evolutionary Anthropology in Leipzig.

Processing of sequenced reads. We used EAGER[63] (version 1.92.50) to process the sequenced reads, using default parameters (see below) for human-originated, UDG-half treated, single-end sequencing data, when processing the UDG-half libraries for all individuals. Specifically, AdapterRemoval was used to trim the sequencing adapters from our reads, with a minimum overlap of 1 bp, and using a minimum base quality of 20 and minimum sequence length of 30 bp. BWA aln (version 0.7.12-r1039, https://sourceforge.net/projects/bio-bwa/files)[64] was used to map the reads to the hs37d5 human reference sequence, with a seed length (-l) of 32, max number of differences (-n) of 0.01 while doing no quality filtering. Duplicate removal was carried out using DeDup v0.12.1. Terminal base deamination damage calculation was done using mapDamage, specifying a length (-l) of 100 bp (Supplementary Table 1).

For downstream analyses, we used bamutils (version 1.0.13, https://github.com/statgen/bamUtil.git) TrimBam to trim two bases at the start and end of all reads. This procedure eliminates the positions that are affected by deamination, thus removing genotyping errors that could arise due to ancient DNA damage.

For three Levänluhta individuals exceeding the threshold coverage of 1% in the preliminary screening, we used the non-UDG treated libraries to confirm the authenticity of the ancient data. For these untreated libraries, two rounds of sequencing were carried out, which were processed using EAGER with the above parameters, but specifying a non-UDG treatment mode and setting the correct sequencing type between the libraries. The merged reads were extracted from the resulting bam files, and merged with the bam file containing reads from the single end sequence run using samtools merge (version 1.3)[65].

The modern Saami genome was generated using Ibis for base calling and an in-house adapter trimming script. The resulting reads were then aligned to the hs37d5 human reference genome using bwa 0.5.9-r16 (parameters -e 20 -o 2 -n 0.01).

Genotyping. We used a custom program (pileupCaller) to genotype the 15 ancient individuals. A pileup file was generated using samtools mpileup with parameters -q 30 -Q 30 -B containing only sites overlapping with our capture panel. From this file, for each individual and each SNP on the 1240K panel, one read covering the SNP was drawn at random, and a pseudohaploid call was made, i.e., the ancient individual was assumed homozygous for the allele on the randomly drawn read for the SNP in question. PileupCaller is available at https://github.com/stschiff/sequenceTools.git.

For the three Levänluhta libraries that did not undergo UDG treatment, we only genotyped transversions, thus eliminating artefacts of post-mortem C- > T damage from further analyses.

The shotgun genome of the modern Saami individual was genotyped using GATK (version 1.3-25-g32cdef9) Unified Genotyper after indel realignment. The variant calls were filtered for variants with a quality score above 30, and a custom script was used to convert the variants into EigenStrat format.

The data were merged with a large dataset consisting of 3871 ancient and modern individuals genotyped on the Human Origins and/or 1240K SNP arrays, using mergeit.

Sex determination. To determine the genetic sex of each ancient individual we calculated the coverage on the autosomes as well as on each sex chromosome. We used a custom script (https://github.com/TCLamnidis/Sex.DetERRmine) for the calculation of each relative coverage as well as their associated error bars (Supplementary Figure 1, Supplementary Note 3 for more information on error calculation). Females are expected to have an x-rate of 1 and a y-rate of 0, while males are expected to have both x- and y-rate of 0.5 (ref. 49).

Authentication. We first confirmed that the deamination pattern at the terminal bases of DNA reads were characteristic of ancient DNA (1.04–4.5% for non-UDG libraries, and 4.7–9.5% for non-UDG libraries, see Supplementary Table 1), using mapDamage (version 2.0.6)[66]. We performed a number of different tests to ensure

the authenticity of our ancient data. For male individuals, we investigated polymorphisms on the X chromosome[27] using the ANGSD software package (version 0.910)[67]. This revealed robust contamination estimates for 2 male Bolshoy individuals, and 1 male Chalmny-Varre individual. All of these were below 1.6% contamination (Table 1). For the female individuals from these two sites, we note that they are projected close to the males in PCA space (Fig. 2a, Supplementary Figure 3), suggesting limited effects of potential contamination. In addition, we generated a PMD-filtered dataset for all individuals using pmdtools (version 0.60)[30]. PMD-filtering was done using a reference genome masked for all positions on the 1240K capture panel, to avoid systematic allelic biases on the analysed SNP positions. We set a pmd-threshold of 3, which, according to the original publication[30], effectively eliminates potential modern contaminants based on the absence of base modifications consistent with deamination.

To provide a more quantitative estimate of possible contamination in females, we used the ContamMix program (version 1.0-10)[29] for estimating mitochondrial contamination. We extracted reads mapping to the mitochondrial reference for each of the ancient individuals using samtools (version 1.3)[65]. We then generated a mitochondrial consensus sequence for each of the ancient individuals using Geneious (version 10.0.9, http://www.geneious.com,[68]), and calling N for all sites with a coverage lower than 5. Finally, all mitochondrial reads were aligned to their respective consensus sequence, using bwa aln (version 0.7.12-r1039)[64] with a maximum number of differences in the seed (-k) set to 5 and the maximum number of differences (-n) to 10, and bwa samse (version 0.7.12-r1039)[64]. A multiple alignment of the consensus sequence and a reference set of 311 mitochondrial genomes[69] was generated, using mafft (version v7.305b)[70–72] with the --auto parameter. The read alignment, as well as the multiple alignment of the consensus and the 311 reference mitochondrial genomes were then provided to ContamMix. We report here the a posteriori mode of contamination, along with the upper and lower bounds of the 95% posterior interval (Table 1).

For additional authentication, we ran supervised ADMIXTURE[28] (version 1.3.0) for all samples using the six present-day populations (Atayal, French, Kalash, Karitiana, Mbuti and Papuan) as defined genetic clusters, to locate any large differences in genetic clustering among individuals from the same site (Supplementary Figure 2). We tested the power of this method to detect contamination and find that it can detect contamination that is distantly related to the ancestries present within the test individuals already at rates of 5–8%, but lacks the power to identify contamination closely related to the test individuals (see Supplementary Note 2). We did not observe significant differences (within our resolution) in the ancestry patterns between the ancient individuals from the same site, with the exception of Levänluhta, where the individual sample JK2065 seems to derive from a different ancestry. We therefore assigned it a separate population label, Levänluhta_B in this study.

Finally, using smartpca, we projected PMD-filtered and non-filtered datasets on the same set of principal components constructed on modern European populations, to ensure that the ancient individuals remain projected in the roughly equivalent positions regardless of PMD-filtering. This was possible for all samples with UDG-half treatment, except for the individuals from Levänluhta, which represented too little damage for the PMD-filtering to be applied. Regarding this site, we therefore relied on the non-UDG libraries (using transversions only) that were generated for the three individuals used in the main analysis. We found that within expected noise due to a low number of SNPs, all samples show consistency between the filtered and non-filtered datasets, suggesting a low amount of contamination in all of the samples (Supplementary Figure 3a, b). Four additional individuals from Levänluhta were excluded from the main analysis and from this authentication test because of low coverage (< 15,000 covered SNPs) and lack of non-UDG libraries.

F statistics. All programmes used for calculating F statistics in this study can be found as part of the Admixtools package (https://github.com/DReichLab/AdmixTools)[2,42].

We used qp3Pop (version 412) for all F_3 calculations.

F_4 statistics were calculated using qpDstat (version 711), and qpAdm (version 632)[2] was used to estimate mixture proportions using the following: Sources (Left Populations): Nganasan; WHG; EHG; Yamnaya_Samara; LBK_EN. Outgroups (Right Populations): OG1: Mbuti; CHG; Israel_Natufian; Onge; Villabruna; Ami; Mixe. OG2: Mbuti; CHG; Onge; Villabruna; Ami; Mixe. OG3: Mixe; CHG; Israel_Natufian; Villabruna; Onge; Ami. OG4: Mbuti; Israel_Natufian; Onge; Villabruna; Ami; Mixe. OG5: Mbuti; Samara_HG; CHG; Israel_Natufian; Villabruna; Ami.

To ensure the outgroup sets had enough power to distinguish the ancestries present in the sources, we ran qpWave (version 410) using only the sources as left populations and each outgroup set as rights. All such qpWave runs were consistent only with maximum rank, meaning all outgroup sets had enough power to distinguish between the five different sources. All qpWave and qpAdm models were run using the option allsnps: YES. When the five-way admixture models provided by qpAdm had p-values above 0.05, but included infeasible mixture proportions and one of the sources was assigned a negative mixture proportion, we ran the model again with that source was excluded. For each Test population, if outgroup set OG1 did not produce a working full model ($p < 0.05$), we tried alternative outgroup sets with one right population removed. This resulted in outgroup sets OG2-4. In the case of Levänluhta, multiple outgroup sets produced working models, which are listed in Supplementary Data 4. The admixture model needing the minimum number of sources while still providing feasible admixture proportions is always shown. In the case of PWC from Sweden where none of the outgroup sets OG1-4 produced a working model, a revised set of right populations was used (OG5) which includes Samara_HG to provide more power to distinguish hunter-gatherer ancestries. We preferred models with OG1-4 over OG5 in general, because OG5 contains more ancient genomes with potential biases in the right populations, which more often causes failing models for modern Test populations. The excluded sources in the minimal models were specified as N/A (Supplementary Data 4). If either Yamnaya or EHG could be dropped (as is the case for Levänluhta), we show the model which is more consistent with previous publications[3,7,8,45] in Fig. 4, but show both models in Supplementary Data 4.

Principal component analysis. We used smartpca (version #16000)[73] (https://github.com/DReichLab/EIG) to carry out Principal Component Analysis (PCA), using the lsqproject: YES and shrinkmode: YES parameters.

For the Eurasian PCA (Fig. 2a), the following populations were used to construct principal components: Abkhasian, Adygei, Albanian, Altaian, Ami, Armenian, Atayal, Avar.SG, Azeri_WGA, Balkar, Balochi, Basque, BedouinA, BedouinB, Belarusian, Borneo, Brahui, Bulgarian, Buryat.SG, Cambodian, Canary_Islanders, Chechen, Chuvash, Croatian, Cypriot, Czech, Dai, Daur, Dolgan, Druze, English, Estonian, Even, Finnish, French, Georgian, Greek, Han, Hazara, Hezhen, Hungarian, Icelandic, Iranian, Italian_North, Italian_South, Japanese, Jew_Ashkenazi, Jew_Georgian, Jew_Iranian, Jew_Iraqi, Jew_Libyan, Jew_Moroccan, Jew_Tunisian, Jew_Turkish, Jew_Yemenite, Jordanian, Kalash, Kalmyk, Kinh, Korean, Kumyk, Kurd_WGA, Kyrgyz, Lahu, Lebanese, Lezgin, Lithuanian, Makrani, Mala, Maltese, Mansi, Miao, Mongola, Mordovian, Naxi, Nganasan, Nogai, North_Ossetian.DG, Norwegian, Orcadian, Oroqen, Palestinian, Pathan, Russian, Saami.DG, Saami_WGA, Sardinian, Saudi, Scottish, Selkup, Semende, She, Sherpa.DG, Sicilian, Spanish, Spanish_North, Syrian, Tajik, Thai, Tibetan.DG, Tu, Tubalar, Tujia, Turkish, Turkmen, Tuvinian, Ukrainian, Ulchi, Uygur, Uzbek, Xibo, Yakut, Yi, Yukagir.

For the West Eurasian PCA (Supplementary Figure 3a, b), the following populations were used to construct principal components: Abkhasian, Adygei, Albanian, Armenian, Balkar, Basque, BedouinA, BedouinB, Belarusian, Bulgarian, Canary_Islander, Chechen, Chuvash, Croatian, Cypriot, Czech, Druze, English, Estonian, Finnish, French, Georgian, Greek, Hungarian, Icelandic, Iranian, Italian_North, Italian_South, Jew_Ashkenazi, Jew_Georgian, Jew_Iranian, Jew_Iraqi, Jew_Libyan, Jew_Moroccan, Jew_Tunisian, Jew_Turkish, Jew_Yemenite, Jordanian, Kumyk, Lebanese, Lezgin, Lithuanian, Maltese, Mordovian, North_Ossetian, Norwegian, Orcadian, Palestinian, Polish, Russian, Sardinian, Saudi, Scottish, Sicilian, Spanish, Spanish_North, Syrian, Turkish, Ukrainian.

ADMIXTURE analysis. ADMIXTURE[28] was run with version 1.3.0, following exclusion of variants with minor allele frequency of 0.01 and after LD pruning using plink (version 1.90b3.29)[74] with a window size of 200, a step size of 5 and an R^2 threshold of 0.5 (https://www.genetics.ucla.edu/software/admixture/download.html). Five replicates were run for each K value, with K values ranging between 2 and 15. The populations used were: Ami, Ami.DG, Armenian, Atayal, Atayal.DG, Balochi, Basque, BedouinB, Belarusian, Brahmin_Tiwari, Brahui, Chuvash, Croatian, Cypriot, Czech, English, Estonian, Even, Finnish, Finnish.DG, French, Greek, GujaratiB, Hadza, Han, Hungarian, Icelandic, Kalash, Karitiana, Lithuanian, Makrani, Mala, Mansi, Mansi.DG, Mari.SG, Mbuti, Mbuti.DG, Mixe, Mordovian, Nganasan, Norwegian, Onge, Orcadian, Papuan, Pima, Russian, Saami.DG, ModernSaami, Sardinian, Scottish, Selkup, Spanish, Ukrainian, Ulchi, Yoruba, ALPC_Hungary_MN, Baalberge_MN, Baltic_BA, Baltic_CCC, Baltic_CWC, Baltic_LN, BolshoyOleniOstrov, Bu_kk_Culture_Hungary_MN, ChalmnyVarre, CHG, EHG, Esperstedt_MN, Ganj_Dareh_Iran_Neolithic, Hungary_MN, Hungary_Neolithic, Iran_Chalcolithic, JK2065, Koros_Hungary_EN, Kunda, Latvia_HG3, Latvia_MN1, LBK_EN, LBK_Hungary_EN, Levanluhta, Narva, PWC_Sweden_NHG.SG, Scandinavia_LNBA, SHG, Sweden_HG.SG, TRB, Ukraine_HG1, Ukraine_N1, WHG, Yamnaya_Samara.

We find that K = 11 results in the lowest Cross-Validation error, as shown in Supplementary Figure 4b.

Y-chromosomal haplotyping. We assigned ancient males to Y haplogroups using the yHaplo program (https://github.com/23andMe/yhaplo)[75]. In short, this program provides an automated search through the Y haplogroup tree (as provided within yHaplo, as accessed from ISOGG on 04 Jan 2016) from the root to the downstream branch based on the presence of derived alleles and assigns the most downstream haplogroup with derived alleles. For about 15,000 Y-chromosomal SNPs present both in our capture panel and in two published datasets[76,77], we randomly sampled a single base and used it as a haploid genotype. We used a custom script to convert EigenStrat genotypes to the yHaplo format. We report the haplogroup assigned furthest downstream provided by the program (Table 1). We also manually checked derived status and absence of mutations defining the designated haplogroup because missing information might lead to a premature stop in its automated search.

Mitochondrial haplotyping. We imported the trimmed mitochondrial reads for each individual with mapping quality >30 into Geneious (version 10.0.9, https://www.geneious.com)[68] and reassembled these reads to the reference genome RSRS[78], using the Geneious mapper, with medium sensitivity and 5 iterations. We used the in-built automated variant caller within Geneious to find mitochondrial polymorphisms with a minimum coverage of 3 and a minimum Variant Frequency of 0.67. The resulting variants were exported to Excel and manually compared to the SNPs reported in the online mtDNA phylogeny (mtDNA tree Build 17, 18 Feb 2016, http://www.phylotree.org/). Nucleotide positions 309.1 C(C), 315.1C, AC indels at 515-522, 16182C, 16183C, 16193.1C(C) and 16519 were masked and not included in our haplotype calls.

Phenotypic SNPs. We used samtools mpileup (version 1.3)[65], filtering for map- (-Q) and base- (-q) quality of 30, deactivating per-Base Alignment Quality (-B), on the trimmed bam files, to generate a pileup of reads mapping to a set of 43 phenotypic SNPs[4,40,41,79] that are part of our genome capture panel. A custom python script was used to parse the pileup into a table containing the number of reads supporting each allele (Supplementary Data 2).

Code availability. All software first described in this study is freely available from online repositories. Sex.DetERRmine: https://github.com/TCLamnidis/Sex.DetERRmine

ContaminateGenotypes: https://github.com/TCLamnidis/ContaminateGenotypes

Data availability

The raw sequence data of the 16 modern and ancient individuals presented in this paper are deposited at the European Nucleotide Archive (http://www.ebi.ac.uk/ena). The study accession is PRJEB29360. A reporting Summary for this Article is available as a Supplementary Information file. The source data underlying all main and Supplementary Figures are provided as a Source Data file.

Received: 8 December 2017 Accepted: 31 October 2018
Published online: 27 November 2018

References

1. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
2. Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
3. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
4. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
5. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
6. Broushaki, F. et al. Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499–503 (2016).
7. Jones, E. R. et al. The Neolithic transition in the Baltic was not driven by admixture with early European farmers. *Curr. Biol.* **27**, 576–582 (2017).
8. Mittnik, A. et al. The genetic prehistory of the Baltic Sea region. *Nat. Commun.* **9**, 442 (2018).
9. Salmela, E. et al. Genome-wide analysis of single nucleotide polymorphisms uncovers population structure in Northern Europe. *PLoS ONE* **3**, e3519 (2008).
10. Huyghe, J. R. et al. A genome-wide analysis of population structure in the Finnish Saami with implications for genetic association studies. *Eur. J. Hum. Genet.* **19**, 347–352 (2011).
11. Lappalainen, T. et al. Regional differences among the Finns: a Y-chromosomal perspective. *Gene* **376**, 207–215 (2006).
12. Kerminen, S. et al. Fine-Scale Genetic Structure in Finland. *G3* **7**, 3459–3468 (2017).
13. Takala, H. O. *The Ristola Site In Lahti And The Earliest Postglacial Settlement Of South Finland* (Lahti City Museum, 2004).
14. Halinen, P. *Prehistoric Hunters of Northernmost Lapland* (Tiedekirja, 2005).
15. Rankama, T. & Kankaanpää, J. in *Early Economy and Settlement in Northern Europe. Pioneering, Resource Use, Coping with Change*Vol. 3 (ed. Blankholm, H. P) 139–167 (Equinox, 2018).
16. Aikio, A. An essay on Saami ethnolinguistic prehistory. *Suom.-Ugr. Seura. Toim.=Mémoires De. la Société Finno-Ougr.* **266**, 63–117 (2012).
17. Aikio, A. *The Saami Loanwords In Finnish and Karelian* (University of Oulu, 2009).
18. Frog, M. & Saarikivi, J. De situ linguarum fennicarum aetatis ferreae, Pars I. *RMN Newsletter* **9**, 64–115 (2014).
19. Rahkonen, P. Onomasticon of Levänluhta and Käldamäki region. *Suomalais-Ugrilaisen Seuran Aikakauskirja* **96**, 287–316 (2017).
20. Häkkinen, J. Jatkuvuusperustelut ja saamelaisen kielen leviäminen (OSA 2). *Muinaistutkija* **2010**, 51–64 (2010).
21. Hausen, R. *Registrum Ecclesiae Aboensis Eller Åbo Domkyrkas Svartbok Med Tillägg Ur Skoklosters Codex Aboensis* (Kansallisarkisto, 1890).
22. Taavitsainen, J.-P. in *The Oxford Handbook of the Archaeology and Anthropology of Hunter-Gatherers* (eds Cummings, V., Jordan, P. & Zvelebil, M.) Ch. 51 (Oxford University Press, 2014).
23. Lang, V. *Läänemeresoome tulemised* Vol. 28 (University of Tartu Publisher, 2018).
24. Der Sarkissian, C. et al. Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of North East Europe. *PLoS Genet.* **9**, e1003296 (2013).
25. Wessman, A. et al. Hidden and remote: new perspectives on the people in the Levänluhta Water Burial, Western Finland (c.ad 300–800). *Eur. J. Archaeol.* **21**, 431–454 (2018).
26. Wessman, A. L. A place of punishment, sacrifice or just a common cemetery? *Fennosc. Archaeol.* **XXVI**, 81–105 (2009).
27. Rasmussen, M. et al. An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98 (2011).
28. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
29. Fu, Q. et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
30. Skoglund, P. et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* **111**, 2229–2234 (2014).
31. Jones, E. R. et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
32. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
33. Raghavan, M. et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
34. Derenko, M. et al. Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia. *PLoS One* **5**, e15214 (2010).
35. Malyarchuk, B., Derenko, M., Denisova, G. & Kravtsova, O. Mitogenomic diversity in Tatars from the Volga-Ural region of Russia. *Mol. Biol. Evol.* **27**, 2220–2226 (2010).
36. Ilumäe, A.-M. et al. Human Y chromosome haplogroup N: a non-trivial time-resolved phylogeography that cuts across language families. *Am. J. Hum. Genet.* **99**, 163–173 (2016).
37. Fondevila, M. et al. Revision of the SNPforID 34-plex forensic ancestry test: assay enhancements, standard reference sample genotypes and extended population studies. *Forensic Sci. Int. Genet* **7**, 63–74 (2013).
38. Kimura, R. et al. A common variation in EDAR is a genetic determinant of shovel-shaped incisors. *Am. J. Hum. Genet.* **85**, 528–535 (2009).
39. Tan, J. et al. The adaptive variant EDARV370A is associated with straight hair in East Asians. *Hum. Genet.* **132**, 1187–1191 (2013).
40. Fumagalli, M. et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* **349**, 1343–1347 (2015).
41. Ameur, A. et al. Genetic adaptation of fatty-acid metabolism: a human-specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids. *Am. J. Hum. Genet.* **90**, 809–820 (2012).
42. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
43. Lahermo, P. et al. Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. *Eur. J. Hum. Genet.* **7**, 447–458 (1999).
44. Csányi, B. et al. Y-chromosome analysis of ancient Hungarian and two modern Hungarian-speaking populations from the Carpathian Basin. *Ann. Hum. Genet.* **72**, 519–534 (2008).
45. Saag, L. et al. Extensive farming in Estonia started through a sex-biased migration from the Steppe. *Curr. Biol.* **27**, 2185–2193 (2017).
46. Loh, P.-R. et al. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233–1254 (2013).
47. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423 (2005).
48. Lipson, M. et al. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017).
49. Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
50. Murashkin, A. I., Kolpakov, E. M., Shumkin, V. Y., Khartanovich, V. I. & Moiseyev, V. G. Kola Oleneostrovskiy Grave Field: A Unique Burial Site In The European Arctic. In *Proc. of the Finnish-Russian Archaeological Symposium* (eds Ulino, P. & Nordqvist, K.) (Tiedekirja, 2016).
51. Carpelan, C. in *Inari–Aanaar. Inarin historia jääkaudesta nykypäivään* (ed. Veli-Pekka, L.) 28–95 (Inarin Kunta, 2003).

52. Carpelan, C. in *Early in the North. Iskos 13* Vol. 5 (ed. Lavento, M.) 30–36 (Finnish Antiquarian Society & Archaeological Society of Finland, 2004).
53. Šumkin, V. J. On the ethnogenesis of the Sami: an archaeological view. *Acta Boreal.* **7**, 3–20 (1990).
54. Honkola, T. et al. Cultural and climatic changes shape the evolutionary history of the Uralic languages. *J. Evol. Biol.* **26**, 1244–1253 (2013).
55. Moiseyev, V. G. & Khartanovich, V. I. Early Metal Age crania from Bolshoy Oleniy Island, Barents Sea. *Archaeol., Ethnol. Anthropol. Eurasia* **40**, 145–154 (2012).
56. Reimer, P. J. et al. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).
57. Ramsey, C. B. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).
58. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci.* USA **110**, 15758–15763 (2013).
59. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448 (2010).
60. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130624 (2015).
61. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
62. Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA.* **110**, 2223–2227 (2013).
63. Peltzer, A. et al. EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
64. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
65. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
66. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
67. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinforma.* **15**, 356 (2014).
68. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
69. Green, R. E. et al. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426 (2008).
70. Katoh, K. & Toh, H. PartTree: an algorithm to build an approximate tree from a large number of unaligned sequences. *Bioinformatics* **23**, 372–374 (2007).
71. Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059–3066 (2002).
72. Katoh, K., Kuma, K.-I., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* **33**, 511–518 (2005).
73. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
74. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
75. David Poznik, G. Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. Preprint at https://www.biorxiv.org/content/early/2016/11/19/088716 (2016).
76. Karmin, M. et al. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* **25**, 459–466 (2015).
77. 1000 Genomes Project Consortium. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
78. Behar, D. M. et al. A 'Copernican' reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
79. Walsh, S. et al. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet* **7**, 98–115 (2013).

Acknowledgements

We thank everyone who contributed to the archaeological excavations. We thank Mikko Putkonen for his notable efforts on early methodological testing and information provided for the Levänluhta samples. We thank Cosimo Posth for carrying out laboratory work, the lab technicians involved in this project, and the sequencing team at the Max Planck Institute for the Science of Human History. We would like to also thank the sequencing team at Max Planck Institute of Evolutionary Anthropology for the sequencing of the modern Saami genome. This project was funded by Emil Aaltonen Foundation, Jane and Aatos Erkko Foundation, the Kone Foundation, Ella and Georg Ehrnrooth Foundation, Jenny and Antti Wihuri Foundation, The Russian State Task for VIGG (AAAA-A16-116111610171-1) and RCMG, the Academy of Finland (grant number: 133056), and the Max Planck Society.

Author contributions

S.S., J.Kr. and W.H. supervised the study. A.Wes., V.M., V.K., A.S., P.O., O.B., W.H. and J.Kr. assembled the collection of archaeological samples. A.S. and S.P. collected the modern Saami sample. K.M. performed laboratory work. K.M. and T.C.L. supervised ancient DNA sequencing and post-sequencing bioinformatics for the ancient individuals. A.Wei. performed the laboratory work of the modern Saami genome. M.O. carried out the processing of the sequenced reads and generating the genotypes of the modern Saami genome. J.Ke. and S.P. supervised sequencing of the modern Saami and post-sequencing bioinformatics. T.C.L., K.M., E.S., C.J. and S.S. analysed genetic data. T.C.L., K.M., E.S., W.H., J.Kr. and S.S. wrote the manuscript with additional input from all other co-authors.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-07483-5.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

5. Manuscript B

Insights into South American population history from ancient genomes from Tierra del Fuego.

Thiseas C. Lamnidis, Zuzana Faltyskova, Christiana Lyn Scheib, Yali Xue, Hannes Schroeder, Susana Morano, Alfredo Prieto, Yolanda Espinosa-Parrilla, Carles Lalueza Fox, Maru Mormina, Richard Durbin, Toomas Kivisild, Stephan Schiffels

Abstract

The American supercontinent was initially populated in multiple waves via the Beringian land bridge. Central and South American populations have been shown to be derived from the first wave of ancestry into the supercontinent. In recent years, the population history of South America has been shown to be more complex than previously thought, entailing multiple waves of ancestry migration from North to South America as well as population replacement events. One of these waves into South America entails ancestry related to that seen in the California Channel Islands, which was detected in the Central Andes 4.2 kya. Here, we present genomic data from 13 ancient individuals from Tierra del Fuego and Patagonia, and additional sequencing data generated for one previously-published individual. When analysing this data together with previously published genomes from Fuego-Patagonia, we find evidence for an introduction of California-Channel-Island-related ancestry into the Kaweskar and Yamana populations between 5,000 and 1,100 years BP. Demographic modelling suggests that ancient Fuego-Patagonian individuals have a generalised affinity to present day South Native American populations.

Introduction

Humans are thought to have first arrived in the Americas through the Beringian land bridge (Flegontov et al., 2019; Llamas et al., 2016; Pedersen et al., 2016; Potter et al., 2018; Raghavan et al., 2015), with evidence for multiple waves of arrival. More specifically, Native American populations were shown to be derived from three independent lines of ancestry. All Central and South American populations, as well as some North American ones, descend from the first ancestry line, termed “First Americans” in (Reich et al., 2012). We will use that term here in light of a lack of established alternatives, acknowledging that it is seen as problematic by some indigenous groups and scholars. Among North American populations, some derive ancestry from a second line of ancestry, linked with the spread of Na-Dene speaking-populations, and others from yet a separate ancestry line, linked to the speakers of Eskimo-Aleutian languages (Flegontov et al., 2019; Reich et al., 2012). It has been suggested that these two later ancestry lines have a genetic connection to Palaeo-Eskimos, such as the Saqqaq culture of Greenland (Flegontov et al., 2019). The ancestral population of First Americans is thought to have split from Siberians roughly 16-22kya (Raghavan et al., 2015), while both First Americans and Siberians have been shown to have ancestry related to both East and West Eurasian lineages. They derive their West Eurasian ancestry from a population termed “Ancient

North Eurasians" (ANE), a West Eurasian population related to Eastern European hunter-gatherers and most often proxied by an Upper Palaeolithic Siberian individual dated to 24kya, while their East Eurasian ancestry is derived from a population related to present-day East Asians (Flegontov et al., 2016; Jeong et al., 2019; Raghavan et al., 2014). An additional affinity of Amazonian populations to Andaman islanders has also been reported (Moreno-Mayar, Vinner, et al., 2018; Skoglund et al., 2015), although this finding is controversial (Posth et al., 2018). In the last half of the 2nd millennium CE, following contact with Europeans and the colonial period that ensued, Native American populations in both North and South America were further shaped by an extreme population bottleneck, as well as admixture with European and/or African populations. The signatures of colonial-time admixture can be very pronounced in present-day Native American populations, although the proportion of ancestry from European and/or African sources is variable across and within populations (1000 Genomes Project Consortium et al., 2015) and individuals with no discernable signatures of colonial-time admixture can still be identified in some populations (Mallick et al., 2016).

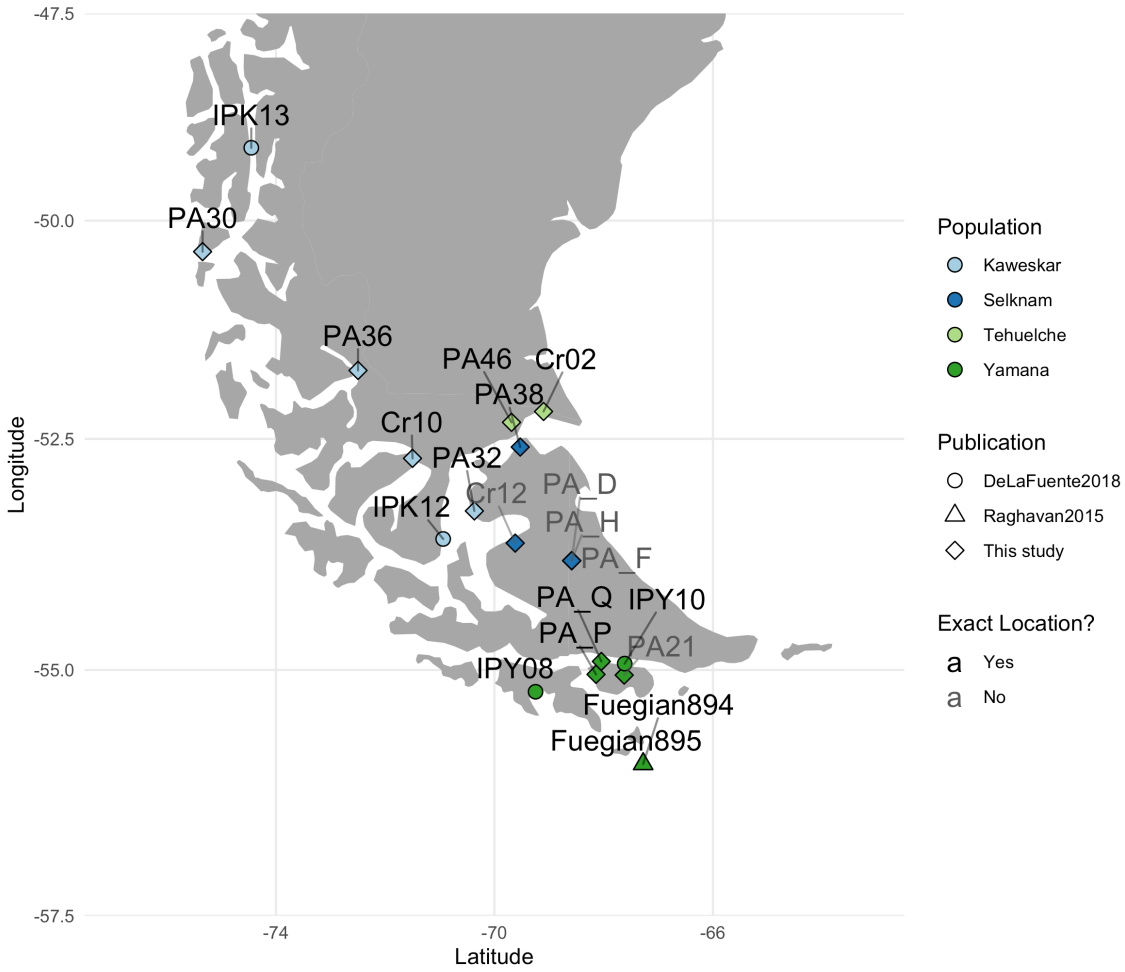
In comparison to North America, South American populations are more homogeneous, since they all descend from the First American ancestry line, with the addition of variable proportions of colonial-time admixture. In recent years, however, the pre-contact population history of South America has been shown to exhibit more subtleties of population structure than previously thought, entailing multiple lines of ancestry between North and South America as well as population replacement events (Moreno-Mayar, Vinner, et al., 2018; Posth et al., 2018; Scheib et al., 2018). Among Native South American populations today and in the past, three ancestry lines have been identified. One ancestry wave is predominantly observed in South America today, one ancestry line is related to Anzick-1 (an individual from Montana dated to 12.5 kya (Rasmussen et al., 2014)), and a final ancestry wave is related to sampled ancient individuals from the California Channel Islands (dating back to 4.9 kya). Within South America, this final ancestry component has only been observed in the Central Andes roughly 4.2 kyBP, however its true geographical extent cannot be determined given the sparse spatial and temporal sampling of ancient individuals in that study (Posth et al., 2018).

Here, we present genomic data from 13 ancient individuals from Tierra del Fuego and Patagonia (including additional data for one previously-published individual) and collate genetic data from previously-published ancient Fuego-Patagonian individuals (de la Fuente et al., 2018; Moreno-Mayar, Vinner, et al., 2018; Raghavan et al., 2015). We find evidence for introduction of ancestry related to the California Channel Islands in the Kaweskar, Selk'nam and Yamana populations in Tierra del Fuego, thus extending the observed geographic range of this ancestry component. Our data suggest the timing of this introduction to be between 5k and 1.1k BP. We confirm claims for genetic continuity in the last millennium within Fuego-Patagonia (de la Fuente et al., 2018), following the introduction of this ancestry. Our demographic modelling suggests that Fuego-Patagonians have generalised affinity to modern South Native American populations, and that some present-day Southern Native American populations may continue to harbour ancestry related to the California Channel Islands ancient population.

Results

Sample information

A total of seventeen individuals from Tierra del Fuego and Patagonia were processed in clean rooms at the University of Cambridge, and libraries were generated without UDG treatment. These libraries were sequenced to low coverage on a NextSeq, and five were selected for deeper shotgun sequencing at the Sanger Institute. We excluded three of the lower coverage individuals from further analysis due to evidence of mitochondrial contamination exceeding a threshold of 5%. The remaining individuals were merged with a dataset of 5681 previously published ancient and modern individuals genotyped on the 1240k SNP panel (Lindo et al., 2018; Mallick et al., 2016; Mathieson et al., 2015; Posth et al., 2018; Scheib et al., 2018), including data from 7 previously published ancient individuals from Tierra del Fuego (de la Fuente et al., 2018; Raghavan et al., 2015). Population attribution as well as the location (when location information was available) of Fuego-Patagonian individuals analysed in this study can be found in Table 1 and Figure 1 respectively. In summary, our dataset contains individuals from four of the five populations found in Tierra del Fuego and southernmost Patagonia, and spans from 300 to 5000 years BP. Note that the archaeological labels in Table 1 are based on regional overlap with observed ethnic groups at time of first contact and not on archaeological findings.



Figure_Map: A map of samples. The colour of each point corresponds to a cultural assignment based on populations present at the time of first contact. The shape of each point corresponds to the origin of the data. The colour of the label signifies the accuracy of the location, with gray labels corresponding to individuals for which exact coordinates were not available. These individuals have been added to the rough centre of the island they were discovered in (PA21,PA_H, PA_D, PA_F), or on the largest modern settlement on the shore of the bay where the individual was discovered (Cr12). No location information is available for FuegianMA577.

Table 1. Sample information

Sample	Population Attribution	Popgen Label	Age	Mean Fold Coverage	Damage 5', 1st base	Genetic Sex	mtDNA Haplotype	Ychr haplotype	mtDNA contamination [contammix]	Nuclear Contamination	#SNPs for 1240k analysis (Tv only) [max 213,539]	SNPs considered for Rare variants (Tv-only) [343,957,486 max]
Cr10	Kaweskar	Fuegian_Kaweskar	-	**5.6268**	0.0909	M	D4h3a5		0.012 (0.010-0.015)	0.29%	210,764	307,394,805
PA30	Kaweskar	Cueva_Ayayema	4619 ± 35BP	**6.9847**	0.1267	M	B2b	Q-M848	0.010 (0.006-0.015)	0.33%	207,514	343,923,090
PA38	Selk'nam	Fuegian_Selknam	290 ± 31 BP	**2.60**	0.0630	M	D4h3a	Q-M848	0.001 (0.000-0.002)	0.44%	195,568	166,062,674
PA_H (H)	Selk'nam	Fuegian_Selknam	318 ± 17BP	**1.7794**	0.0658	M	D1g	Q-M848	0.008 (0.005-0.011)	1.73%	173,032	89,738,508
PA_Q (Q)	Yamana	Fuegian_Yamana_600BP	657 ± 34 BP	**9.2808**	0.0878	F	C1b	N/A	0.011 (0.008-0.016)	N/A	208,813	327,619,505
Cr02	Tehuelche	Fuegian_Tehuelche	-	0.3042	0.0726	M	C1b		0.014 (0.005-0.029)	0.39%	59,951	Excluded (<1x)
Cr12	Selknam	Fuegian_Selknam	-	0.1263	0.0638	F	D1g	N/A	0.000 (0.000-0.003)	N/A	26,206	Excluded (<1x)
PA_D (D)	Selknam	Fuegian_Selknam	-	0.0933	0.0435	F	D1g	N/A	0.001 (0.000-0.008)	N/A	16,944	Excluded (<1x)
PA_F (F)	Selknam	Fuegian_Selknam	-	0.2254	0.0682	M	D1g		0.001 (0.000-0.007)	0.28%	44,514	Excluded (<1x)
PA_P (P)	Yamana	Fuegian_Yamana_1100BP	1369 ± 16 BP	0.0739	0.1377	F	C1b	N/A	0.009 (0.004-0.015)	N/A	14,736	Excluded (<1x)
PA20	Yamana		659 ± 22 BP	0.1305	0.0603	F	C1b	N/A	0.075 (0.046-0.116)	N/A	27,306	Excluded (<1x)
PA21	Yamana	Fuegian_Yamana_600BP	906 15 BP	0.0815	0.0444	M	C1b		0.044 (0.021-0.088)	13.51%	16,834	Excluded (<1x)
PA32	Kaweskar	Fuegian_Kaweskar	271 ± 23 BP	0.471	0.0219	F	D1g	N/A	0.000 (0.000-0.003)	N/A	77,089	Excluded (<1x)
PA36	Kaweskar	Fuegian_Kaweskar	1623 ± 22 BP	0.244	0.0926	M	D4h3a5		0.012 (0.005-0.021)	-0.84%	46,694	Excluded (<1x)
PA45	Tehuelche		242 ±21 BP	0.0744	0.0369	M	D1g		0.113 (0.064-0.185)	N/A	16,402	Excluded (<1x)

PA46	Tehuelche	Fuegian_Tehuelche	3908 ± 31 BP	0.1674	0.2249	M	B2b		0.009 (0.003-0.017)	3.83%	34,126	Excluded (<1x)
ZN	?		-	0.0073	0.0234	M	D1g		0.135 (0.023-0.974)	N/A	1,648	Excluded (<1x)
IPK12	Kaweskar	Fuegian_Kaweskar_1100BP	1,000 ± 30 BP	7.8	0.1336	F	C1b		0.015 (0.010-0.021)		211,079	338,660,541
IPK13	Kaweskar	Fuegian_Kaweskar_1100BP	1,320 ± 30 BP	3.5	0.0989	M	D1g+16189	Q1a2a1a1-M3	0.005 (0.004-0.008)		206,746	260,272,630
IPY08	Yamana	Fuegian_Yamana_1100BP	-	1.7	0.0850	M	D4h3a5	Q1a2a1a1-M4	0.004 (0.002-0.008)		188,138	133,145,943
IPY10	Yamana	Fuegian_Yamana_1100BP	910 ± 30 BP	9.1	0.0655	M	C1b	Q1a2a1a1-M5	0.003 (0.002-0.005)		211,620	341,480,992
Fuegian894	Yamana	Fuegian894	200 BP	1.0	N/A	F	D4h3a	N/A	N/A		118,130	40,483,796
Fuegian895	Yamana	Fuegian895	200 BP	1.3	0.03	F	C1b	N/A	N/A		125,855	56,718,589
FuegianMA577	Selknam	FuegianMA577	200 BP	1.7	0.05	M	D1g5	N/A	N/A		179,089	104,528,680

Genetic Kinship Analysis

To identify any genetic relatives within our dataset we calculated the pairwise mismatch rate of genotypes (pMMR) among the ancient Fuego-Patagonians (Figure 2). The low pMMR between PA30 and the published data from an individual discovered in Ayayema Cave , is indicative of these two samples being from the same individual. Hence, the sequencing data for this individual was merged (Cueva_Ayayema). We do not identify any genetic relatives among the ancient Fuego-Patagonian individuals. The lower-than-median pMMR observed in the rest of the comparisons shown in Figure 2 is a result of population structure, since the median is calculated from all possible comparisons among ancient Fuego-Patagonians, which also includes comparisons both within and between Fuegian populations.

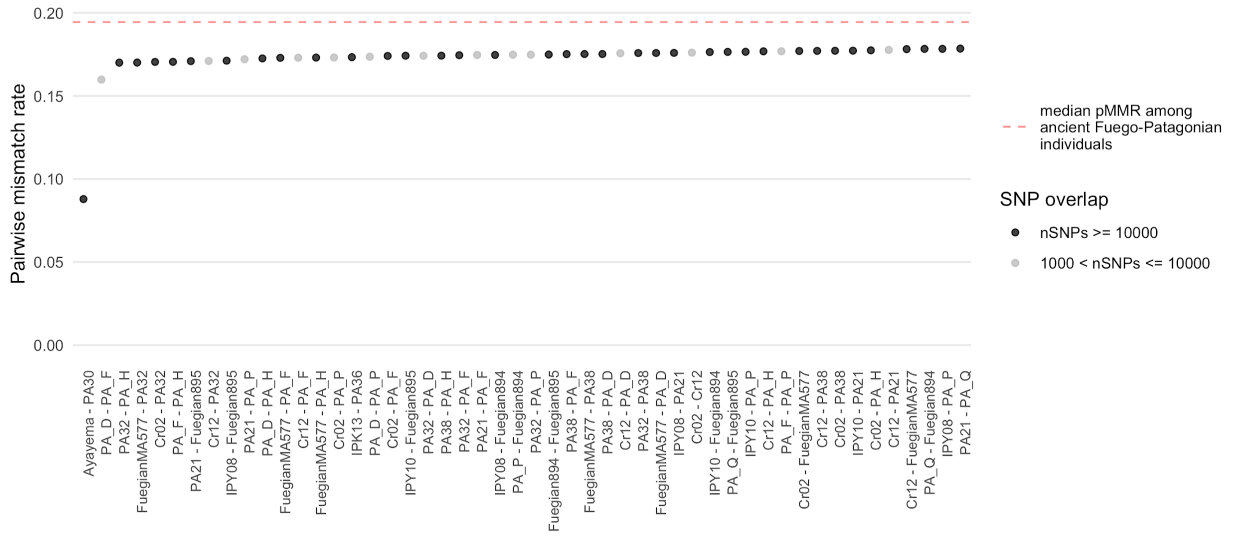


Figure 2: The lowest 50 pairwise mismatch rate of genotypes observed among ancient Fuego-Patagonians. The colour of the points denotes the number of overlapping SNPs in each comparison. The median value of the mismatch rate is shown as a red dashed line.

No evidence of non-American genetic contributions

Radiocarbon dates associated with some of the analysed individuals are post-dating contact between people from the Old World and the peoples of South America. We used principal component analysis (PCA) to project the ancient Fuego-Patagonians to principal components of genetic variation calculated on data from 135 modern worldwide populations to identify any colonial-time genetic contribution from Europe or elsewhere in the ancestry of the studied individuals (Figure 3).

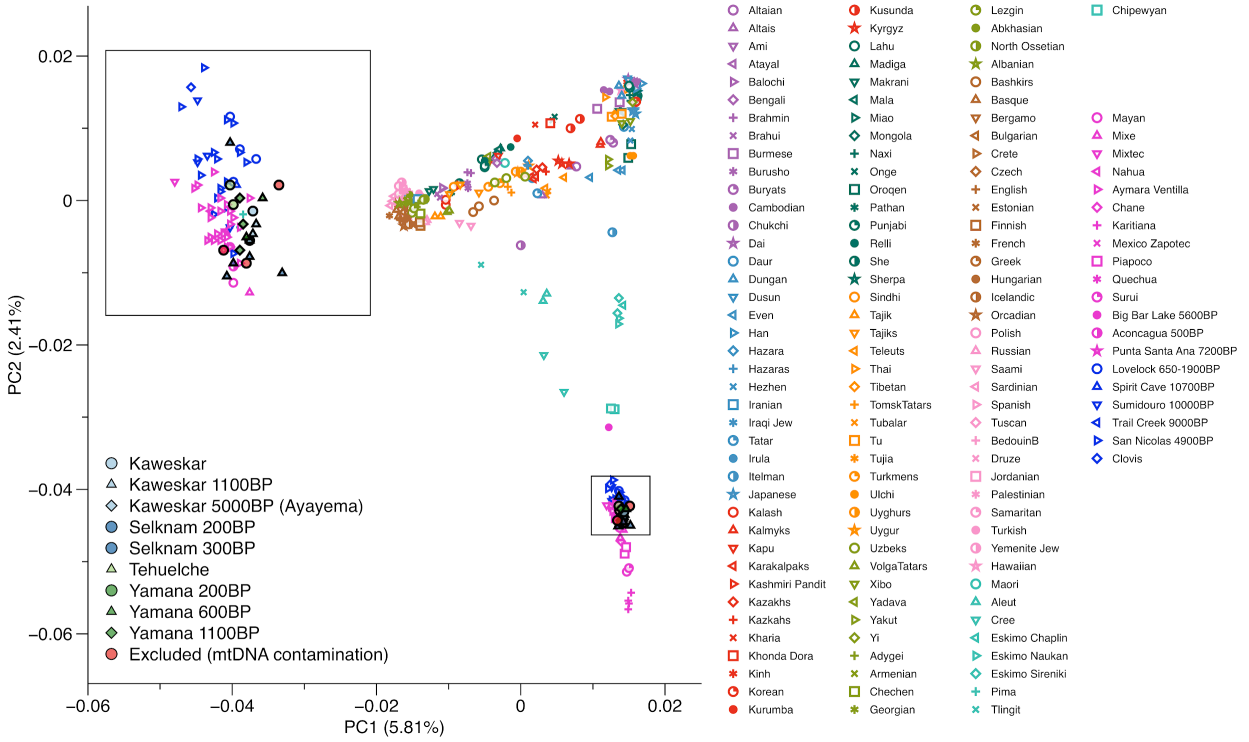


Figure 3: A Principal Components Analysis from 132 Eurasian and Native American populations. Ancient individuals are projected with the “lsqproject: YES” option, and shrinkage is corrected with “shrinkmode: ON”. Among ancient Fuegian groups, colour designated cultural association, while shape designates the different groups within each cultural group. Groupings shown here are consistent with qpWave. Three ancient Fuegians with evidence for mitochondrial contamination that were removed from further analyses are shown in red.

The ancient Fuego-Patagonians are projected within the South American genetic cluster (in magenta). ADMIXTURE analysis reveals a similar pattern, whereby the ancestry of Fuego-Patagonians is modelled entirely by a component that is maximised in Native American populations (Figure 4), thus confirming the lack of colonial ancestry in the studied individuals.

Figure 4: Genetic clustering using ADMIXTURE. The replicate with the lowest CV error is shown for runs with K=2 to K=6.

Population dynamics within Fuego-Patagonian populations

Population grouping of the ancient individuals

Ethnic attribution of ancient remains in the area is based on the geographical ranges of the different Fuego-Patagonian populations. Principal component analysis carried out on the

variation of 11 Native American populations (Figure 5) reveals the different Fuego-Patagonian individuals cluster with modern day Aymara individuals (Lindo et al., 2018). However, this PCA does not reveal clear genetic clusters between Fuego-Patagonians populations, or through time.

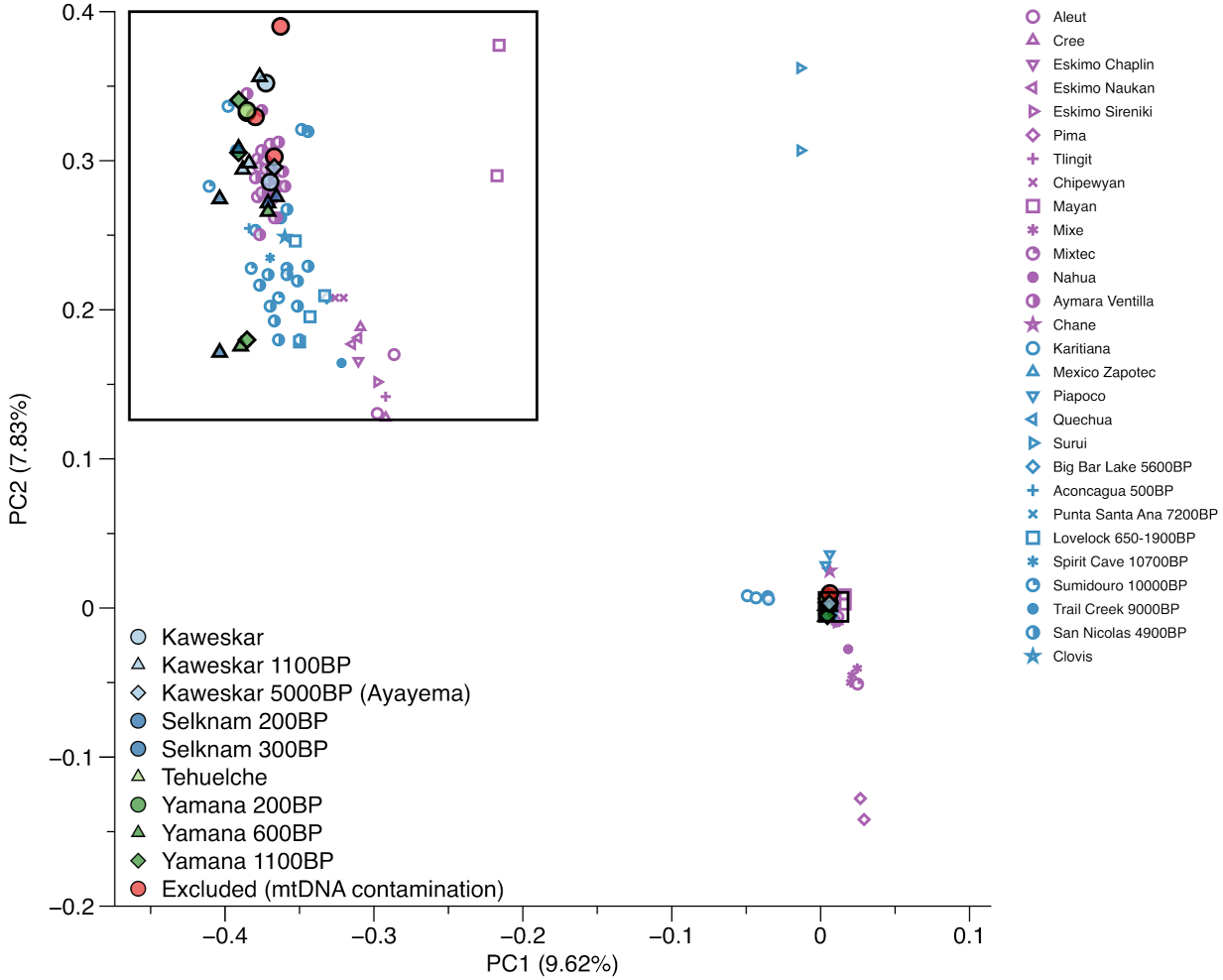


Figure 5: A Principal Components Analysis on variation of 11 Native American populations. Ancient and modern individuals are projected with the "lsqproject: YES" option, and shrinkage is corrected with "shrinkmode: ON". Among ancient Fuegian groups, colour designated cultural association, while shape designates the different groups within each cultural group. Groupings shown here are consistent with qpWave. Three ancient Fuegians with evidence for mitochondrial contamination that were removed from further analyses are shown in red.

Given the temporal span of our dataset, we decided to test for genetic continuity between individuals with the same ethnic attribution. We used *qpWave* to calculate the lines of ancestry required to explain the genetic data of all individuals attributed to each Fuego-Patagonian population. For each Fuego-Patagonian population of size n, we modelled the number of ancestry streams from a set of present-day Native American populations (and Mbuti, a present-day African hunter gatherer group) required to explain n "Left" populations, comprising each ancient individual attributed to this population. If more than one ancestry stream was required, all possible models for $n-1$ Left populations were run, and this process repeated until all n individuals had been assigned to groups that are genetically homogeneous to the limit of our

resolution. Any individuals that were consistent with more than one of the resulting groups (presumably as a result of the low number of covered SNPs) were added to the group that was closest to their C14 date (if one was available). Following this approach, individuals attributed to the Kaweskar were split into three groups; The 5250 cal BP individual (Cueva_Ayayema) composes the first group; one group consists of the two individuals dated to 860-1190 cal BP; and one final group comprising the remaining individuals. Similarly, the individuals attributed to the Yamana were separated into two groups; one comprising two individuals dated between 810-1240 cal BP, and one consisting of the remaining individuals (dated 600-760 cal BP). The individuals attributed to the Selknam and Tehuelche populations were explained as a single ancestry stream respectively, and were not split into subgroups.

Admixture Graph Modelling

To assess the population dynamics in Fuego-Patagonia through time within each of the sampled populations, we reconstructed the admixture graph from (Posth et al., 2018) and used that graph as a scaffold for further analysis. Summary statistics between the model and the data agree, with only 3 out of 2211 tested f-statistics showing deviations between the observed and expected values of the statistic. These deviations were up to 3.2 times the standard error from each other. (worst Z=-3.219, 3 outliers). Each of the seven genetic groups we identified was then added to the scaffold tree while refitting all parameters, either as a result of a simple split from any edge, or as the result of admixture of any two edges in the scaffold. If a simple split from an edge resulted in a worst absolute Z score below 3.5, then the results explaining the group as an admixture edge were not considered, to avoid over-fitting. Among equally well-fitting models, the models that produced no 0 length branches directly upstream or downstream of the added branch were preferred. The best fitting graphs for each Fuego-Patagonian group are shown in Figure 6.

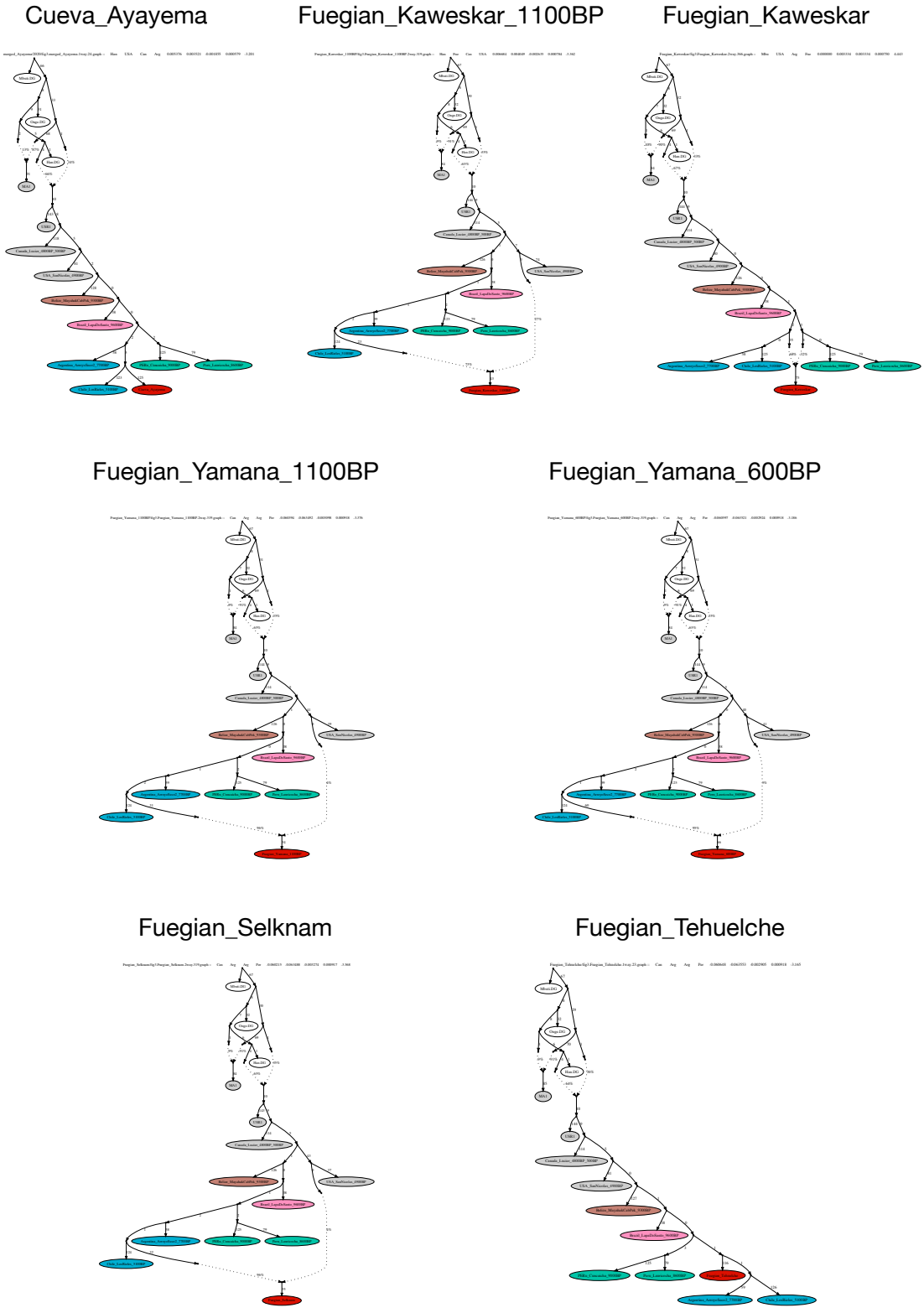


Figure 6: Best fitting qpGraph model for each Fuego-Patagonian group.

Our results show that:

- **Cueva_Ayayema** is consistently modelled (worst Z=3.201, 3 outliers) (Figure 6) as a part of the radiation of populations in Chile and Argentina (shown in blue in the imported image). The only position in this clade that does not produce 0 drift lengths is as a sister clade to the 5100-year-old Chilean population from Los Rieles.
- **Fuegian_Kaweskar_1100BP** can be modelled as having a majority component from the population radiation of Chile and Argentina, and a minor component of 20-30% from an ancestry related to USA_San_Nicholas_4900BP (worst Z=-3.362, 6 outliers).
- **Fuegian_Kaweskar** cannot be successfully modelled within this context (worst Z=4.443, 40+ outliers).
- **Fuegian_Yamana_1100BP** is modelled as having a majority component from the population radiation of Chile and Argentina, and a minor component of 4% from an ancestry related to USA_San_Nicholas_4900BP (worst Z=-3.376, 3-5 outliers).
- **Fuegian_Yamana_600BP** is modelled as having a majority component from the population radiation of Chile and Argentina, and a minor component of 5% from an ancestry related to USA_San_Nicholas_4900BP (worst Z=-3.185, 3 outliers).
- **Fuegian_Tehuelche** is also modelled well as part of the radiation in Chile and Argentina (worst Z=-3.165, 5 outliers). We do not have resolution to confidently infer the exact position of this group within the clade.
- Finally, **Fuegian_Selknam** can be modelled as having a majority component from the population radiation of Chile and Argentina, and a minor component of 4% from an ancestry related to USA_San_Nicholas_4900BP (worst Z=-3.568, 6-9 outliers).

Collectively, these results show a dynamic genetic profile for the Kaweskar in the last 5000 years. By 1100 years BP, a new genetic component related to the earliest sampled ancient individuals from San Nicolas Island (ESN) is identified within the Kaweskar, and the same component is also seen in the Yamana, yet to a smaller extent. Since the Yamana population is inhabiting more southern areas than the Kaweskar, this pattern could be the result of an ancestry wave coming from the North, and which gets diluted as it moves into more southern areas. Concerning the lack of well fitting admixture models for the Fuegian_Kaweskar group, we believe this is an artefact of low SNP overlap between the two low-coverage individuals in this group, which hides structure within the population.

Relationship of ancient Fuego-Patagonians and modern South Americans

To investigate the relationship of ancient South Americans to present-day populations, we developed a new qpGraph to model present-day populations from South America. We model modern South American populations as arising from a single admixture event between Ancient North Eurasian and East Eurasian lineages (worst Z=-3.086, 1 outlier) (Raghavan et al., 2014) (Figure 7).

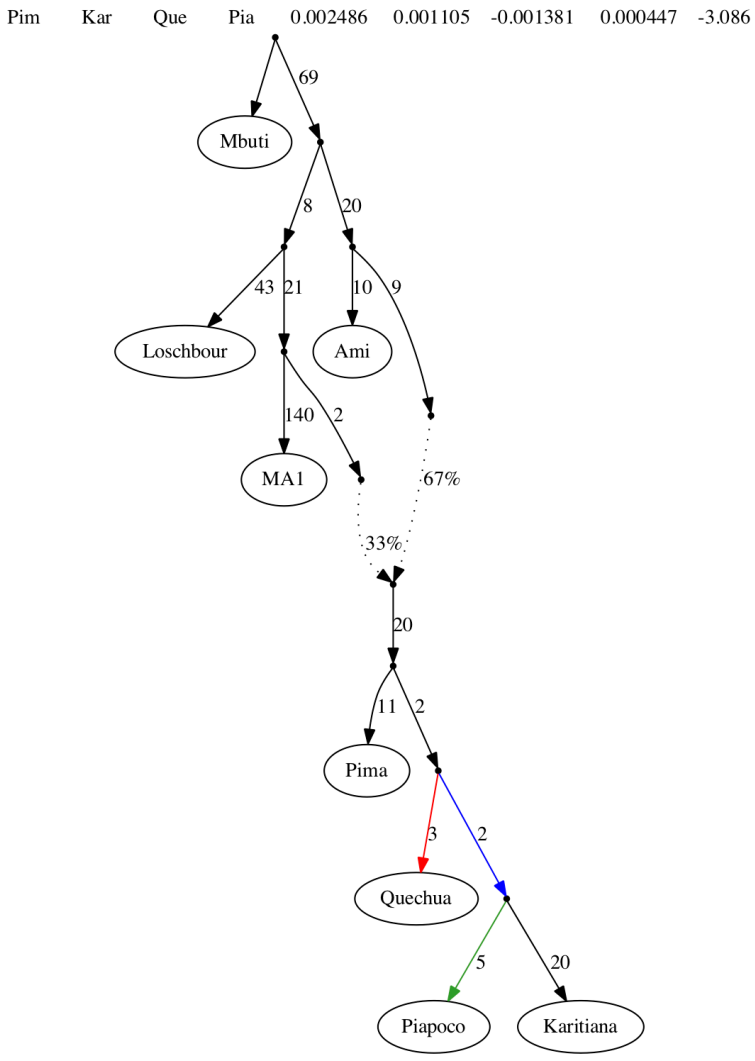


Figure 7: The coloured edges highlight the best-fitting placement position of different ancient populations.

We forced the placement of different Fuego-Patagonian groups as simple clades originating from each edge in Figure 7. Placements that did not create branches with length 0 directly upstream or downstream of the added node were preferred. In the case of Fuegian_Tehuelche, where all models produced edges of length 0, the model with the lowest chi-square score was

selected. This approach resulted in well-fitting models (worst Z-score ≤ 3.5) for almost all groups, with the exception of Fuegian_Kaweskar, who were best modelled as a two-way mixture between groups. Specifically:

- The oldest Kaweskar individual from Ayayema Cave was modelled as a sister clade to Quechua (shown in red in Figure 7) (worst Z-score=-3.049, 1 outlier).
- All Yamana and Selknam, as well as the group Kaweskar_1100BP are best modelled as outgroups to the modern Piapoco and Karitiana populations (blue in Figure 7) (worst Z-score=-3.322 to 2.986, 0-1 outliers)
- The group of ancient Tehuelche is best fitted as splitting off from the ancestor node of all South Americans (parent node of red and blue edges) (worst Z-score=-3.105, 3 outliers). However, in this model, the branch length leading to this group was zero, indicating a lack of lineage-specific drift. This is likely a result of lower coverage in this group compared to the other ancient groups.
- The group "Fuegian_Kaweskar" is best fitted as the result of admixture between two edges, deriving 63% of its ancestry from a Quechua-related lineage (red) and 37% from a Piapoco-related lineage (green).

Adding the different ancient Fuego-Patagonians to this model shows differences between the groups, consistent with the finding using the previous model, where the oldest Kaweskar (Cueva_Ayayama) is the only Kaweskar individual without additional ancestry. Ancient individuals best modelled as deriving part of their ancestry from a population related to the California Channel Islands are consistently modelled here as being more closely related to Piapoco/Karitiana, and less related to Quechua. These results imply that differences in proportions of ancestry related to each of the independent ancestry streams between North and South America are, in part, responsible for the genetic differences observed between present-day South American populations.

Rarecoal

Progressive building of a scaffold tree

The analytical methods we have used thus far are based on differences between populations on specific common markers across the genome. Such methods cannot estimate absolute timing of demographic events, such as populations splits or admixture events. However, given the number of whole-genome data available from South America, in addition to the data generated for this study, it is possible to use rare variation to model population demography. This approach has been used in the past to successfully model population history (Flegontov et al., 2019; Schiffels et al., 2016). To this end, we compiled a dataset including whole-genome data available for present-day and ancient individuals with a specific focus in the Americas. We used *rarecoal* to build a demographic model of worldwide populations with a specific focus on South American populations. We compiled a dataset of variants found in published ancient and modern data from different sources (Mallick et al., 2016; Raghavan et al., 2015), and genotyped all ancient individuals with average genome coverage above 1x using the approach described in (Flegontov et al., 2019). Briefly, each ancient individual was genotyped only in sites that were found to be bi-allelic in our dataset of published individuals and that was covered by at least 3

reads. The genotypes were called by randomly drawing three reads overlapping each variable position and using a majority call to create a single haplotype per ancient individual. This approach has been argued to mitigate coverage-dependent reference bias.

We built the demographic model using *rarecoal* by progressively adding populations and parameters for the model to fit, following (Flegontov et al., 2019). At each stage, the fits of observed and expected Rare Allele Sharing (RAS) were checked and the model optimised. The resulting model includes super-populations for Europeans (EUR, n=33), South-East Asians (SEA, n=21), and Siberians (SIB, n=22), a ghost Ancient North Eurasian (ANE) lineage, as well as three First American populations (Zapotec from Mexico, Piapoco from Colombia, and Quechua from Peru, n=2 for each population).

We began by modelling the split of Southeast Asians (SEA) and Europeans (EUR). We find a well-fitting model with an estimated split time at 48,200 yBP (Figure 8).

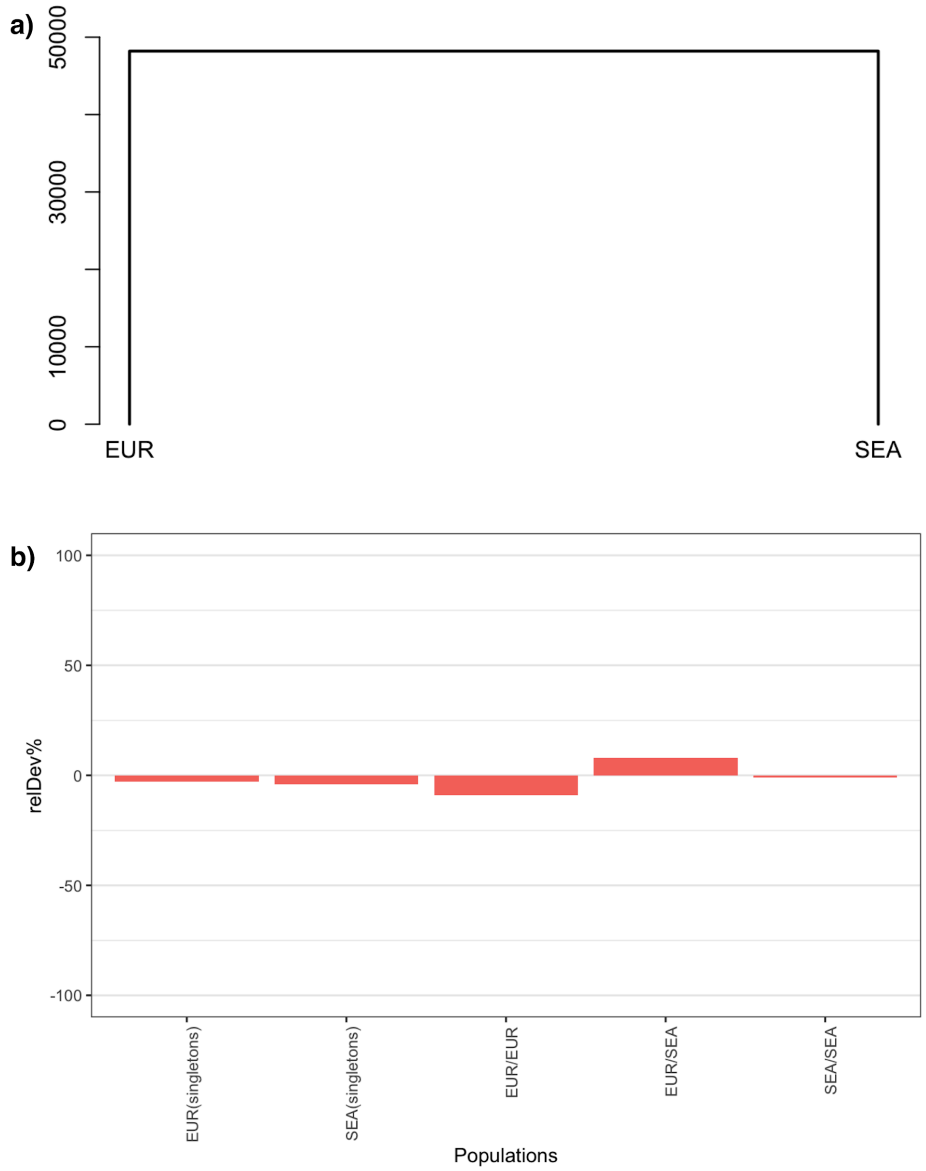


Figure 8: a) Model tree. b) Relative deviation of observed and expected rare allele sharing.

Next, we added a supergroup of all Southern Native Americans (SAM, n=13) to this tree. In accordance with previous research (Flegontov et al., 2019; Patterson et al., 2012; Raghavan et al., 2014), we model this population as deriving part of its ancestry from a ghost population of Ancient North Eurasians (ANE) most closely related to EUR (Figure 9).

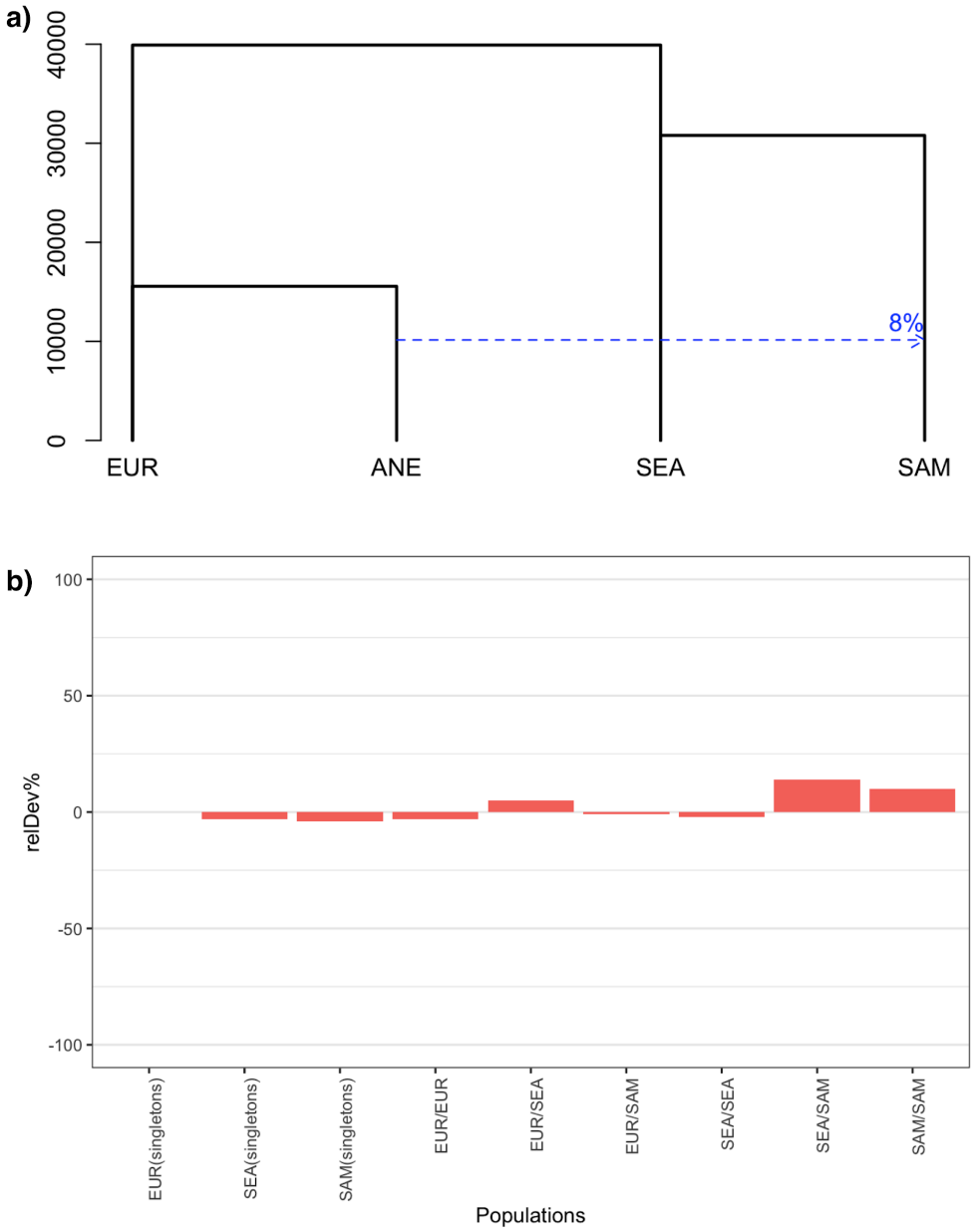


Figure 9: a) Model tree. b) Relative deviation of observed and expected rare allele sharing.

We further extended this model by adding Siberians (SIB) as a sister group to SAM. To optimise the model based on the relative deviations of observed and expected rare allele sharing, we introduced two additional admixture events, one from the EUR population into SIB with an estimated magnitude of 15.2% happening roughly 6,000 yBP, and another from the ANE branch

into the common ancestor of the SIB and SAM branches (Figure 10). The latter admixture edge was fixed at the same time as the population split of SIB and SAM populations (estimated to ~19,000 yBP).

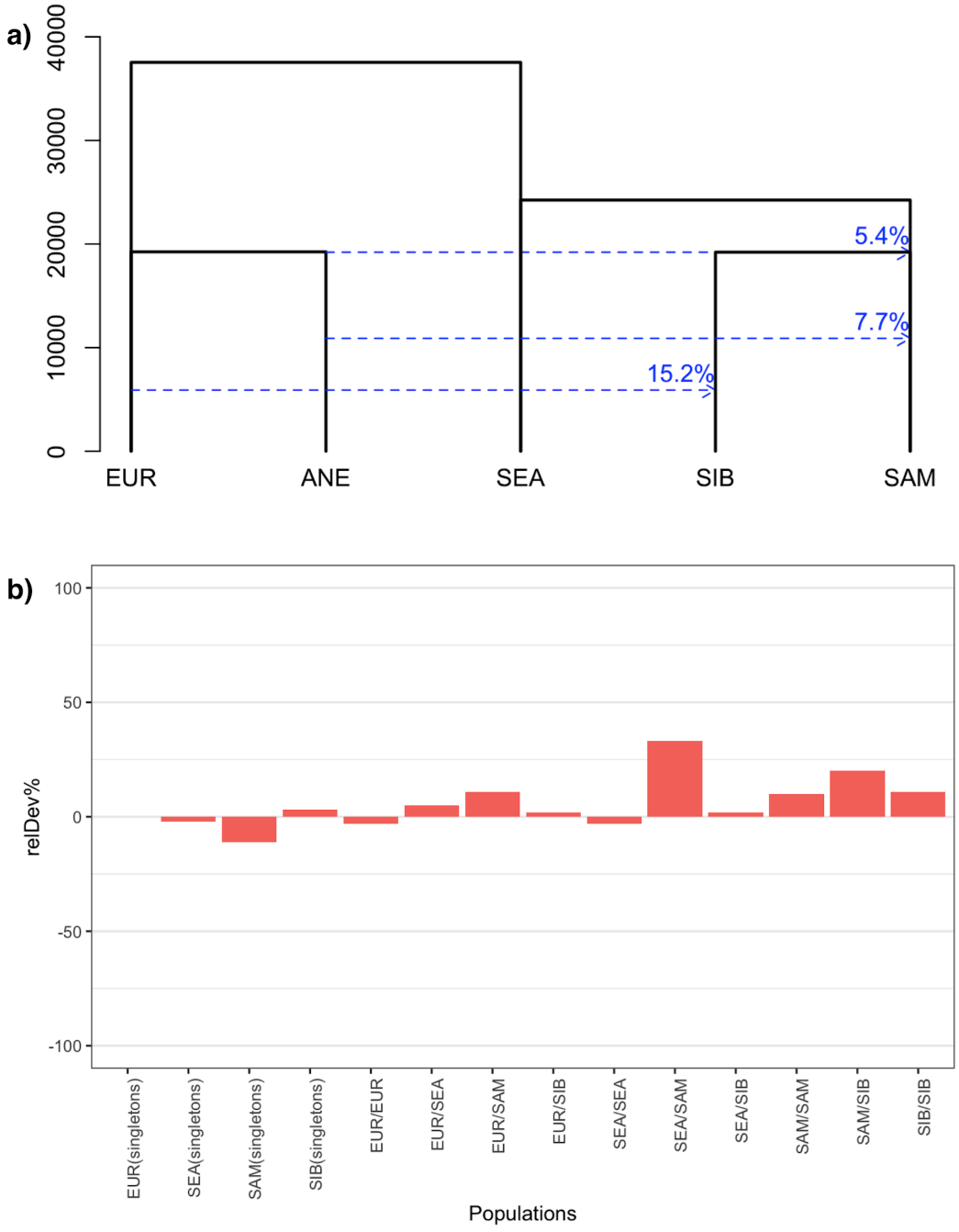


Figure 10: a) Model tree. b) Relative deviation of observed and expected rare allele sharing.

In an attempt to gain more resolution within South America, we began replacing the SAM superpopulation with a set of central and southern Native Americans. We began by replacing the superpopulation with the Zapotec, a Native American population from Mexico (Figure 11).

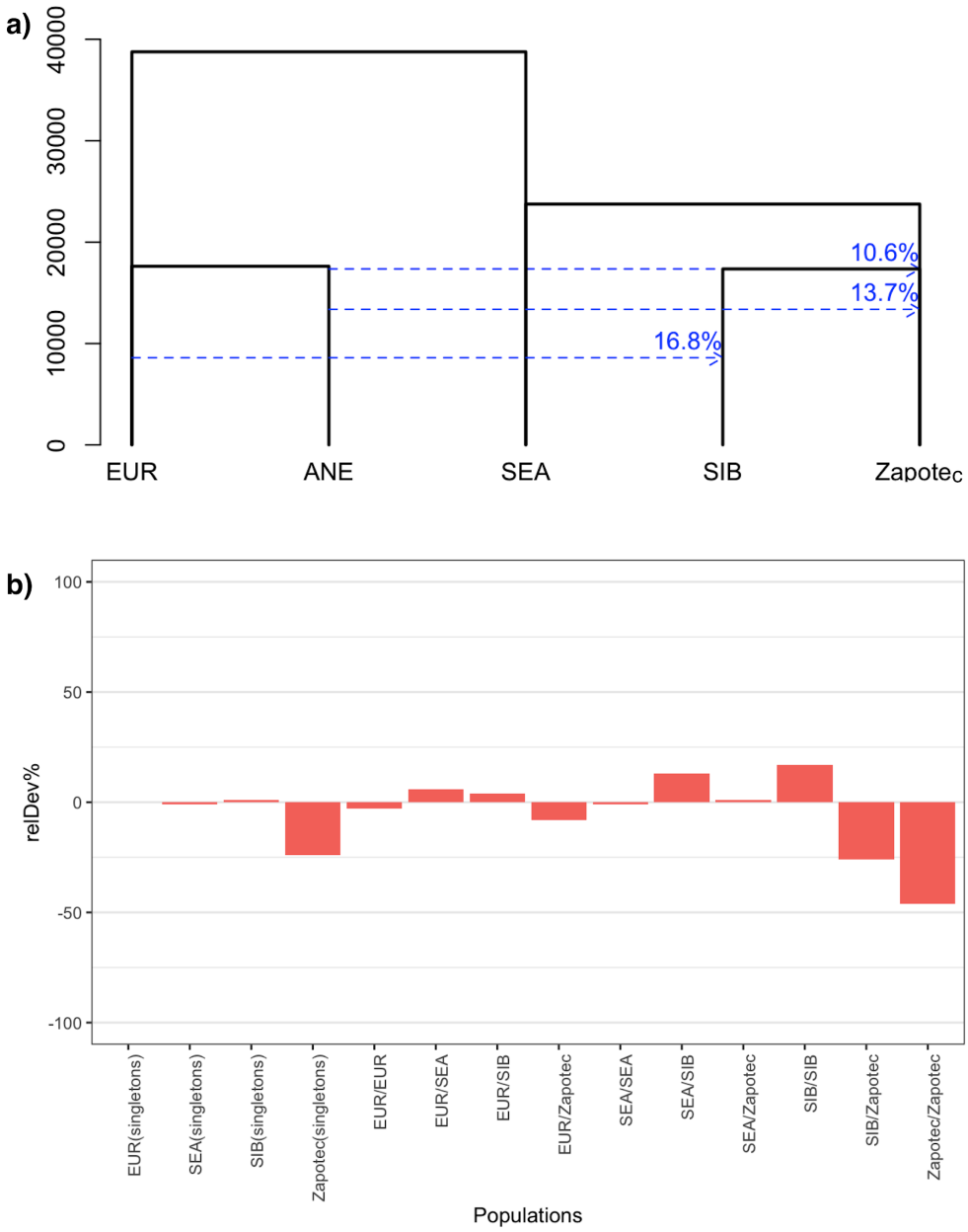


Figure 11: a) Model tree. b) Relative deviation of observed and expected rare allele sharing.

Next, Piapoco, a South American group from Colombia, were added as a sister group to Zapotec (Figure 12). To better model the differences in ANE-related ancestry in SIB and SAM populations as resulting from population structure rather than two serial but independent admixture events, we fixed the timing of the admixture from ANE into the common ancestor of all Native American populations (NAM). For this and all subsequent models, the timing of this admixture edge was set to happen directly after the split of SIB and NAM. In this model, the ancestral admixture from ANE into the SIB/NAM ancestor is estimated at 7.8%, while the subsequent ANE admixture into NAM is estimated at 13.6%.

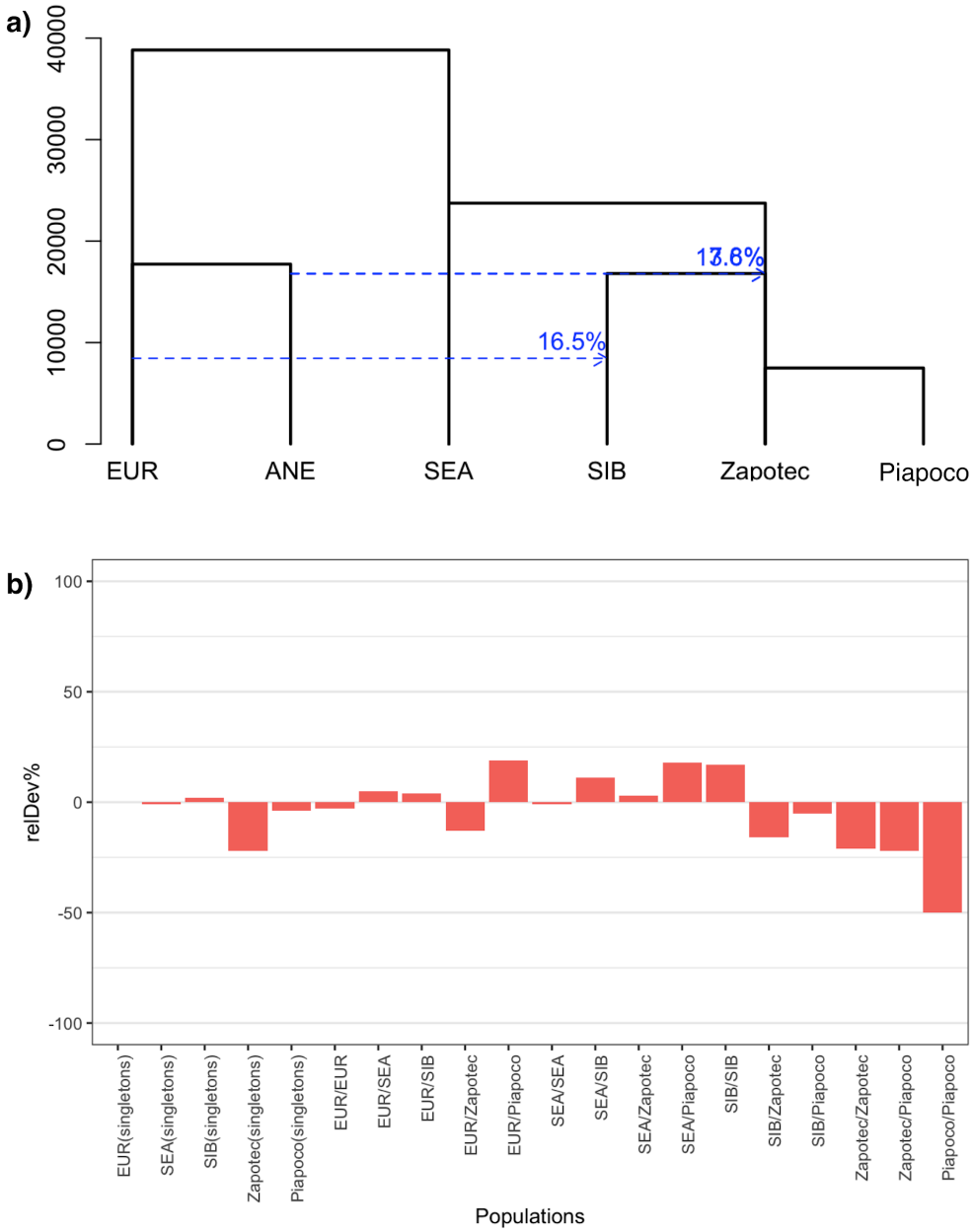


Figure 12: a) Model tree. b) Relative deviation of observed and expected rare allele sharing.

Finally, the Quechua, an Andean population from Peru, were added to the model as a sister clade to the Piapoco. The inferred tree (Figure 13) is largely consistent with published literature, and fits the observed RAS well (Figure 14, no over/under-estimations above 50%, only 1 above 25%). Note that fits with rarecoal are never expected to fit perfectly (like qpGraph), because rarecoal fits thousands of variables with a few dozen parameters at most, while in qpGraph the number of variables (F stats) and the number of parameters are typically more similar.

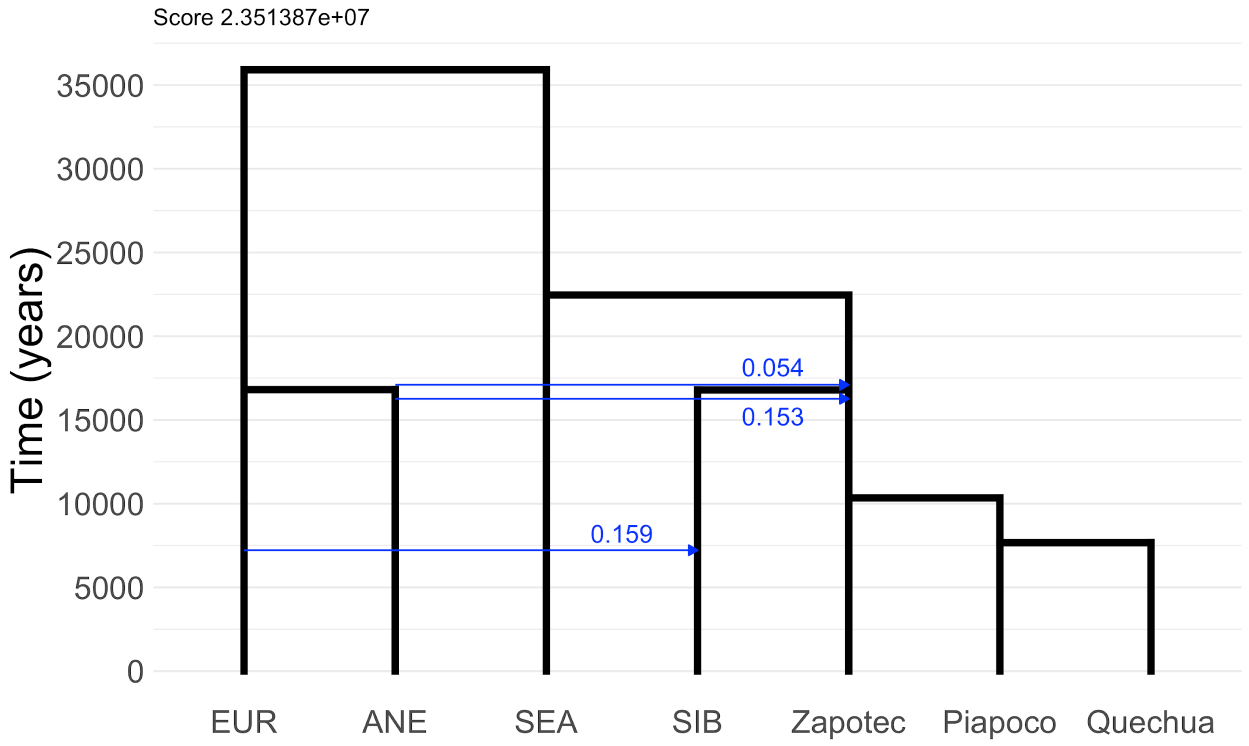


Figure 13: Scaffold tree inferred with rarecoal.

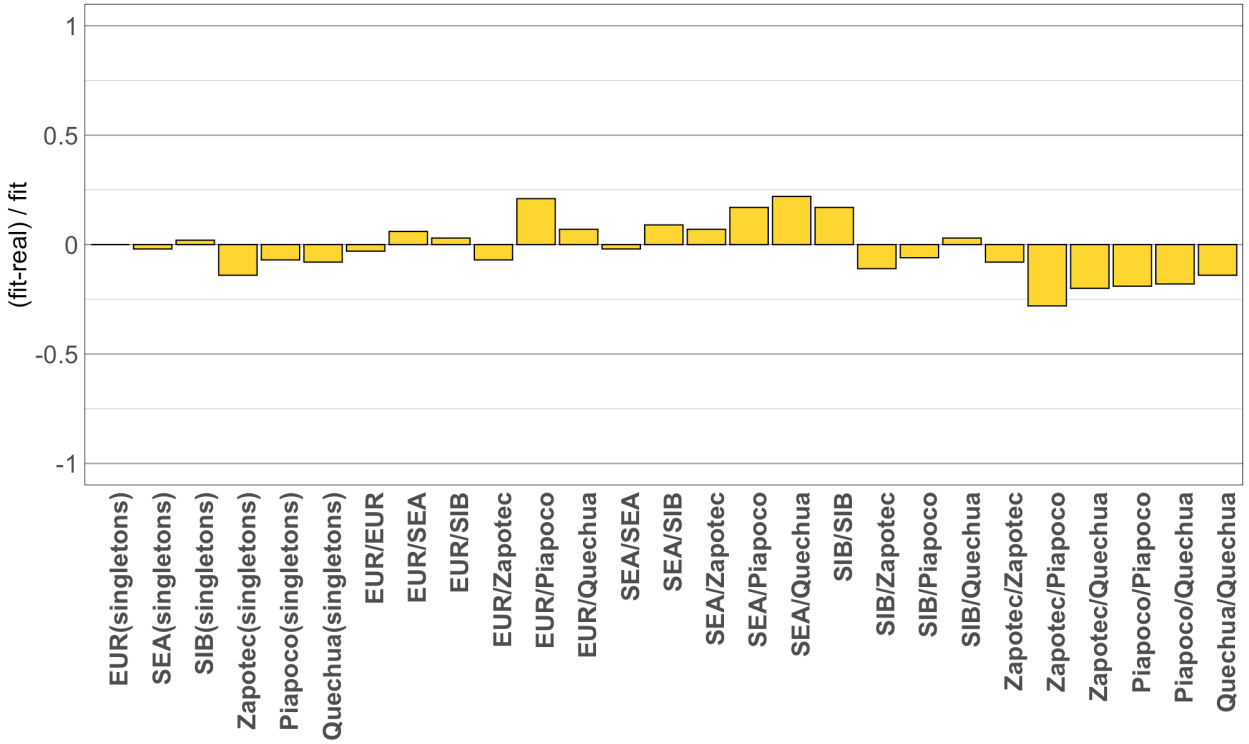


Figure 14: Fit of observed to expected rare allele sharing inferred by rarecoal mcmc. These fits correspond to the tree shown in Figure 13.

Two important features of the resulting tree are the estimated split time between EUR and ANE, and the two admixture events into the Native American (NAM) lineage the timings of which were

fixed at the split time of the SIB and NAM lineages. The inferred split time of EUR and ANE lineages is estimated to be more recent than previously suggested. This is likely a result of using European genomes with ANE-related ancestry to infer the ANE lineage, instead of ancient genomes. The two admixture edges correspond to one admixture into the SIB/NAM ancestral population, and a second admixture in the ancestral NAM population. We believe that this setup better describes a structured common ancestral population with an ANE-related genetic contribution for both the SIB and NAM lineages than a model with an independent admixture edge for each branch would (see also modelling in (Flegontov et al., 2019)).

According to this model, EUR and SEA populations split around 40,000 years ago, followed by a population split between SEA and the ancestral population of both SIB and NAM roughly 22,50 year ago. The SIB and NAM populations subsequently split around 17,000 years ago. The inferred population size for the ancestral NAM following the separation from SIB is 4-5 times lower than the inferred population size of the SIB/NAM ancestral population, a finding that reflects the deep bottleneck experienced by NAM populations. We estimate the split time between South American populations (Piapoco & Quechua) and Central American populations (Zapotec) to 10,350 years ago, while the separation of Quechua and Piapoco is inferred to have taken place 7,650 years ago. The Native American populations used in the model have been previously shown to be unadmixed (Mallick et al., 2016)

Adding ancient individuals

We used *rarecoal find* to investigate the most likely split point for each ancient Fuego-Patagonian individual. *Rarecoal find* will query every branch at multiple equally spaced time points as candidate split points for a given population, and report the parent branch and split time that best fits the data. Ancient branches were added with the respective branch-shortening, according to the age of the sample. For individuals that have not been carbon dated, we fitted a linear regression to predict the carbon date of an individual based on the ancient DNA damage pattern of the first bp in the 5' terminus of DNA fragments. The resulting equation was significant, ($F(1,12)=15.6$, $p<.005$) with an R^2 of 0.53 (Figure 15).

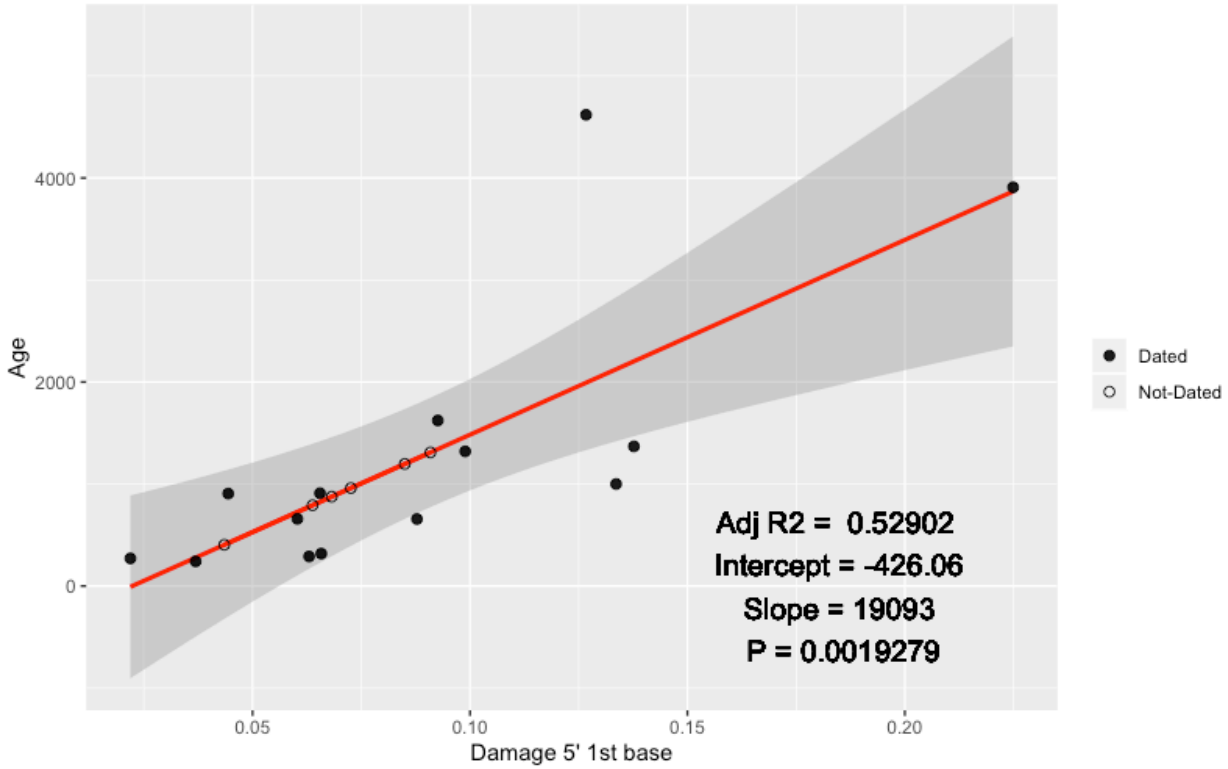


Figure 15

All ancient Fuego-Patagonian genomes were best modelled as branching off from either Quechua (Cr10, Cueva_Ayayema, PA38, Fuegian894) or Piapoco (PA_H, PA_Q, IPK12, IPK13, IPY08, IPY10, Fuegian895, FuegianMA577), at times ranging from 4970-6940 years BP (Figure 16). We followed the same approach to add relevant previously-published ancient genomes from the Americas to the scaffold tree (de la Fuente et al., 2018; Moreno-Mayar, Potter, et al., 2018; Moreno-Mayar, Vinner, et al., 2018; Raghavan et al., 2015). We used *rarecoal fitTable* to investigate the differences between expected and observed RAS between the different ancient individuals and the populations in the scaffold tree. Overall, the inferred placement of the ancient individuals captures the sharing of rare alleles with modern populations very well (Figure 17). Some structure can be observed among the Fuego-Patagonian individuals, however it is not captured well with *rarecoal find*, since it tests for simple splits without any admixture. Most ancient Fuego-Patagonian individuals share most rare alleles with the Quechua. Four individuals (Cr10, Cueva_Ayayema, PA38, PA_H) show roughly equal sharing with both the Quechua and the Piapoco.

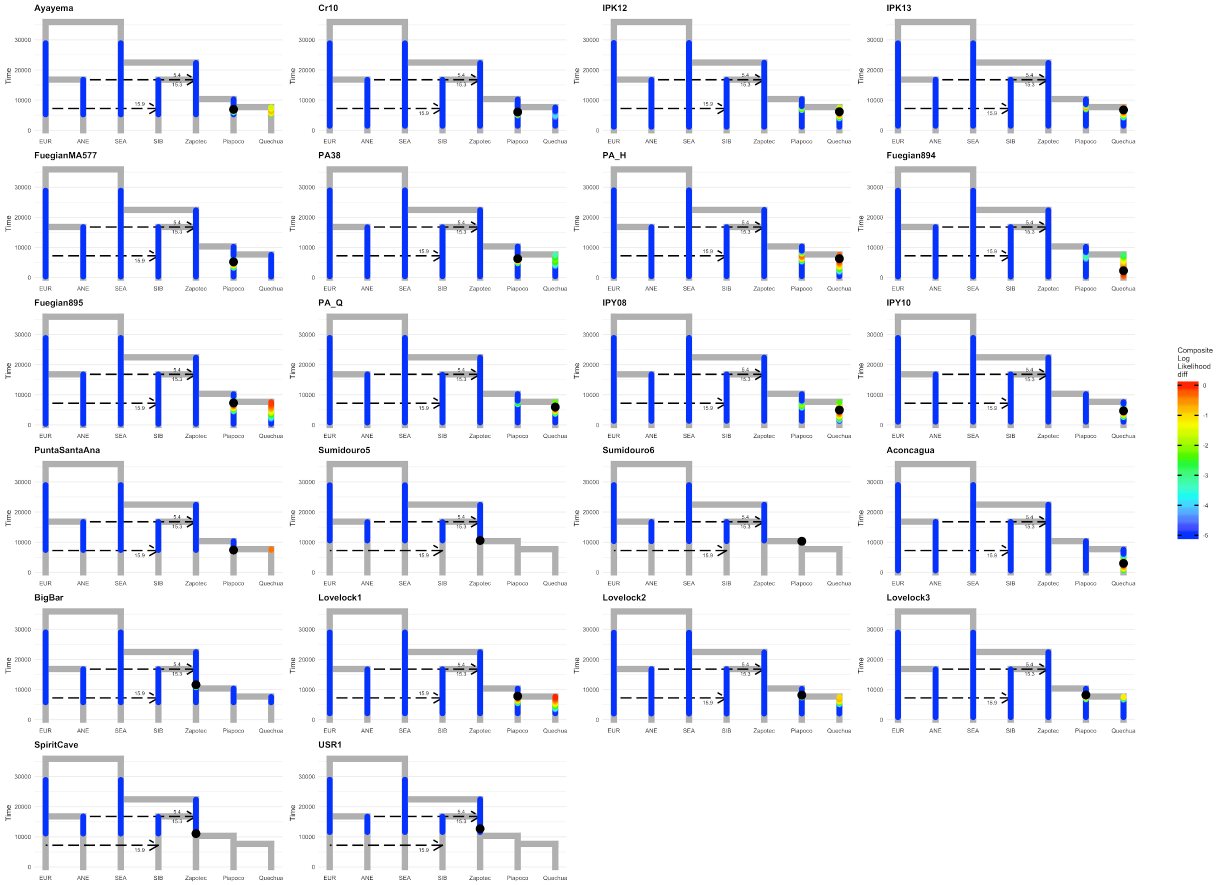


Figure 16: The results of *rarecoal find* for each tested ancient individual. The maximum likelihood branching position for each individual is indicated by the black point. The rest of the tested branching points are coloured based on their composite likelihood difference to the best branching position. Composite likelihood differences larger than or equal to 5 are shown in dark blue.

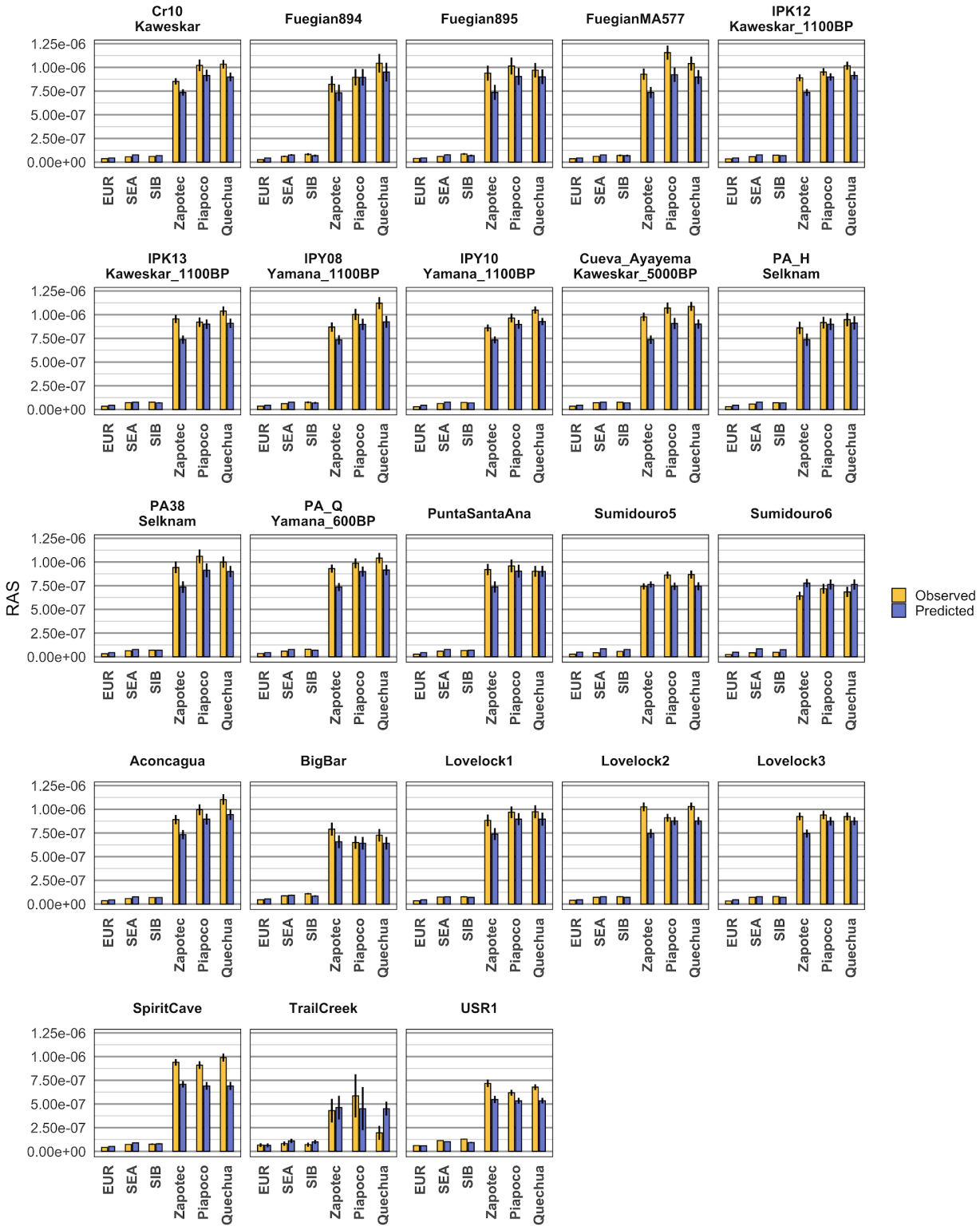


Figure 17

To investigate the confidence interval around the highest likelihood placement inferred by *rarecoal find*, and to correct for the effects of genetic linkage, we followed a procedure to reduce the effective number of sites described in Flegontov et. al 2019. Briefly, we apply a jackknifing

approach (Busing et al., 1999) implemented as part of *freqSum2Histogram* to estimate standard errors around the fraction of sites following each possible sharing pattern across the six (super-)populations in the scaffold tree.

We apply a least-squares fit to infer a scaling factor α for the effective number of sites. We find this factor to be equal to 0.0087 (equivalent to a reducing the effective number of sites 115 times), and apply this scaling to the composite log-likelihood difference to obtain corrected log-likelihood differences (Figure 18). A model with a corrected log-Likelihood difference of -2 is e^2 = 7.39 times less likely than the maximum likelihood model. We used this value (Corrected log-likelihood difference $\geqslant$ -2) as an arbitrary cutoff for the possible placement of the different ancient individuals. We observe that our placement of ancient individuals among the modern Native American populations has limited resolution, with small likelihood differences across all branching positions from the Native American lineages. However, for all Fuego-Patagonian individuals branching positions on the ancestral Native American branch are rejected. Some North American individuals with high-coverage genomes that were placed on the Zapotec branch (SpiritCave, USR1) retain small confidence intervals regardless of scaling. In contrast, lower coverage genomes from North America (Lovelock, BigBar) show very broad confidence intervals, equivalent to generalised affinity towards all modern Native American populations.

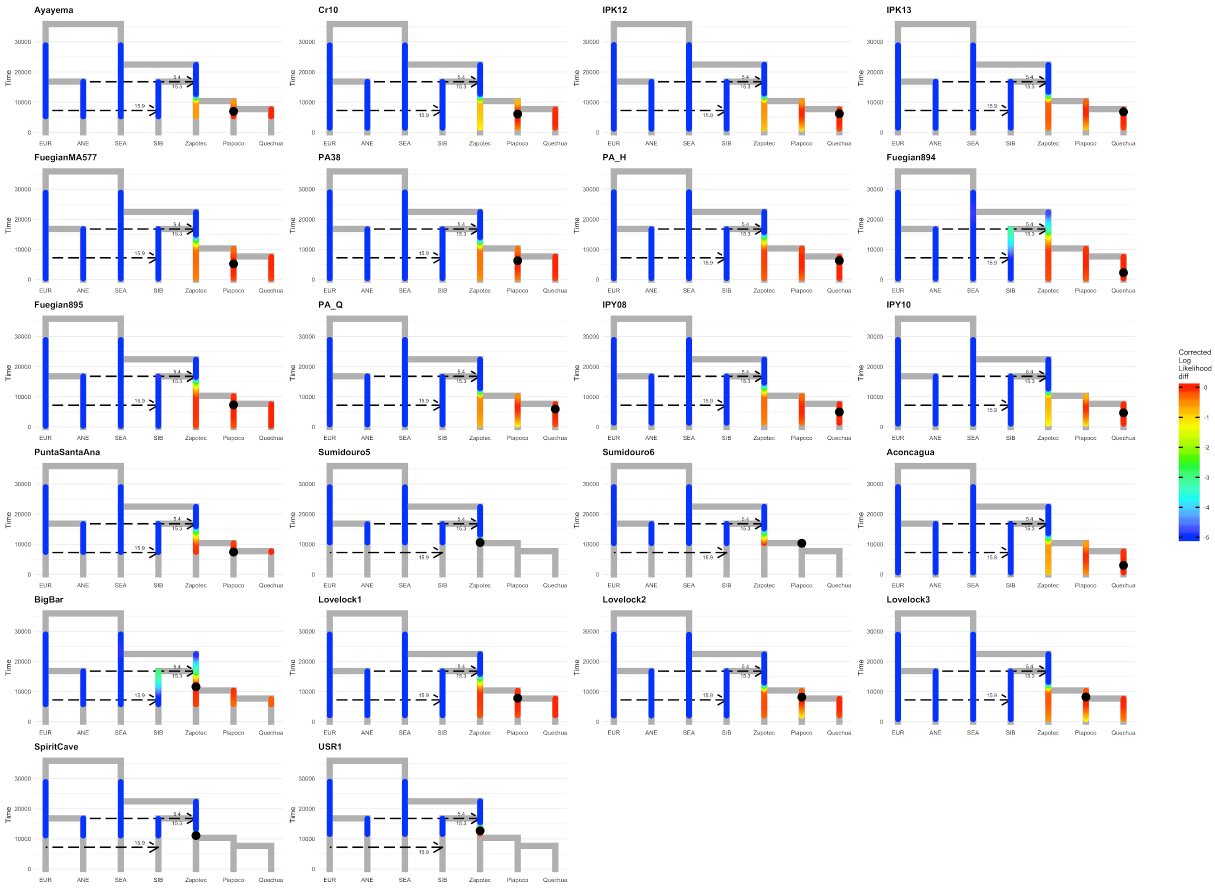


Figure 18: The results of *rarecoal find* for each tested ancient individual. The maximum likelihood branching position for each individual is indicated by the black point. The rest of the tested branching points are coloured based on their composite likelihood difference to the best branching position.

Composite likelihood differences larger than or equal to 5 are shown in dark blue. The effective number of sites for the models shown here has been rescaled by a factor 0.0087.

The scaling factor we estimate is different to and significantly smaller (i.e. rendering scaled results less confident) than that estimated for a similar population phylogeny based on a similar dataset in previously published work (Flegontov et al., 2019). To investigate the difference this discrepancy makes in the confidence interval around each placement, we also scaled the effective number of sites using the previously reported scaling factor (0.055, equivalent to reducing the effective number of sites 18 times) (Figure 19). With this scaling factor, fewer branching positions remain below the arbitrary corrected log likelihood difference we have set. Notably, branching positions on the Zapotec branch for most ancient Fuego-Patagonian individuals can be rejected. This is not the case for individuals Fuegian894, Fuegian895, FuegianMA577, IPY08, and PA_H, who have the lowest coverage among ancient Fuego-Patagonians (mean genome coverage between 1x and 1.7x).

A consistent signal we see in the ancient Fuegian data is the relatively deep maximum likelihood branch-point of most individuals, including relatively high coverage samples. This deep placement then also results in the broad affinity we see with other branches, leading to Zapotec and Piapoco. Such a deep placement is not observed with an ancient Argentinian mummy of moderate coverage, which is more safely placed onto the Quechua branch, at a point much more recent than for most ancient Fuegians. This suggests that the broad branch affinity seen in Fuegians may reflect their specific population history rather than coverage or resolution.

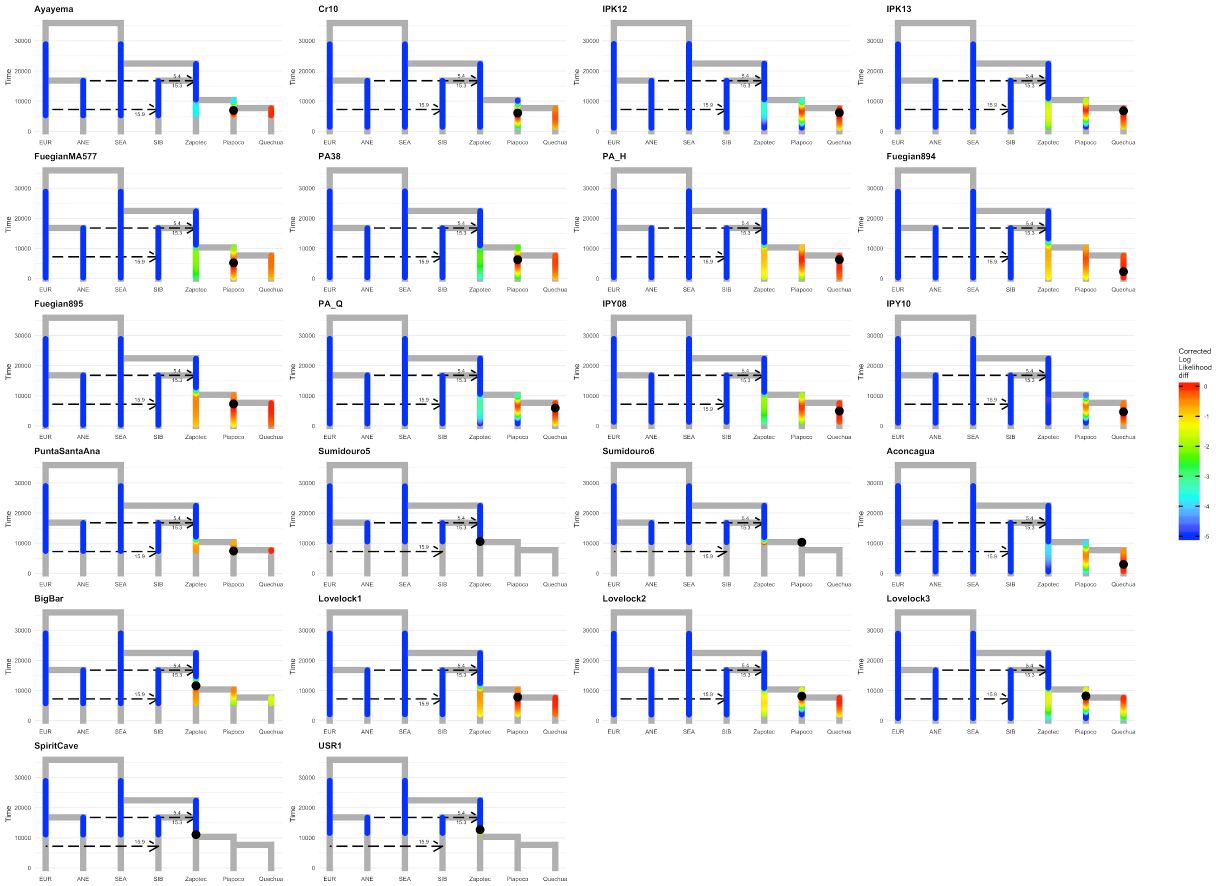


Figure 19: The results of *rarecoal find* for each tested ancient individual. The maximum likelihood branching position for each individual is indicated by the black point. The rest of the tested branching points are coloured based on their composite likelihood difference to the best branching position. Composite likelihood differences larger than or equal to 5 are shown in dark blue. The effective number of sites for the models shown here has been rescaled by a factor 0.055.

Discussion

Our results show (the first evidence for) population changes within Tierra del Fuego during the last 4.5 ky. Consistent with the genetic shift in the rest of South America, we detect an influx of ancestry related to the Early San Nicholas in the Kaweskar. This genetic shift happens at some time between 5 and 1 kya, as evidenced by the lack of this component in the 5kya individual from Ayayema Cave, and detection of the component in all more recent Kaweskar individuals. Similar genetic affinities are observed in the Yamana and Selknam, albeit to a lower extent. This geographical pattern could arise if the ESN-like ancestry is entering Tierra del Fuego from the northwest (where the Kaweskar resided) and gets progressively diluted as it travels southward and eastward. The lack of this ancestry in the samples Tehuelches could be due to temporal factors, since the one individual for which radiocarbon dating was possible has been dated to 4285 calBP, thus possibly predating the arrival of ESN-related ancestry in the region. It should be noted that a similar genetic component has previously been identified in Central Adean populations (Posth et al., 2018). Our finding suggests that ESN-related ancestry was more prevalent than previous evidence suggested.

In terms of their relationship to modern South American populations, our findings show that the ancient populations without detectable ESN-related ancestry are also less related to the Piapoco and Karitiana, sharing no specific drift with these populations that is not also shared with Quechua. One possible explanation for this result is that Karitiana and Piapoco also harbour some ESN-related ancestry.

Our demographic model based on the sharing of rare genetic variation captures the observed patterns of sharing very well. We infer a split time of 10.3 kya between Central and South American populations, as well as a population split between Quechua and Piapoco roughly 7.6 kya. Placement of ancient individuals onto this demographic model reveals limited resolution to distinguish placement among South American populations. There are two possible explanations for these results; Limited resolution, and population history. The Native American populations used in demographic modelling consist of 2 individuals each. This could cause somewhat common variation within/among these Native American populations to pass our rarity cutoff of being observed between 2 and 4 times in the dataset. More common variants are more likely to be shared between Native American populations, and therefore be less informative about ancient branch placement, thus driving broader affinities. This limitation will partly influence the confidence intervals around each placement. However, it is also likely that ancient Fuego-Patagonians exhibit generalised affinity to the modern South American populations as a result of population history events within South America. Indeed, we confidently infer the branching position of Aconcagua (an Argentinian mummy dated to ~500 BP) to be roughly 3 kya on the

Quechua branch, with limited "spillover" to the Piapoco branch (Figure 19). Limitations in resolution caused by the population sizes of the modern reference populations would be affecting this individual as well and hence these cannot be solely responsible for the generalised affinity observed. This difference in confidence intervals also cannot be explained by coverage differences between the ancient individuals, since Aconcagua is of lower mean genomic coverage than Fuego-Patagonian individuals who exhibit wider confidence intervals (e.g. Ayayema). The effect of smaller reference population size on these inferences should be investigated in the future, when more present-day high-coverage genomes from the Americas become available.

Methods

DNA extraction and library preparation

DNA for seventeen ancient individuals was extracted in the clean-room facilities of the University of Cambridge Department of Archaeology Ancient DNA Laboratory.

Sequencing

The whole genomes for individuals Cr10, PA30, PA38, PA_H, and PA_Q were sequenced on a HiSeq2500 by the Wellcome Sanger Institute, using a paired-end 75 cycle kit.
Additional low-coverage whole genomes were generated at the University of Cambridge Department of Biochemistry DNA Sequencing Facility on NextSeq500 single-end 75 cycle kit.

Processing of sequenced reads

Sequenced reads were processed using the EAGER pipeline(Peltzer et al., 2016) (version 1.92.55), using default parameters. Removal of adapters from the sequences was done using AdapterRemoval, using a minimum base quality of 20, a minimum sequence length of 30 and a minimum overlap of 1bp. We used BWA aln (version 0.7.12-r1039, https://sourceforge.net/projects/bio-bwa/files) to map the sequenced reads to the human reference sequence (hs37d5). For this step, the maximum number of differences (-n) was set to 0.01 and seed length (-l) was set to 32. DeDup (v0.12.1) was used to remove PCR duplicates. MapDamage was used to calculate read-terminus deamination rate, specifying a length (-l) of 100 bp.

Genotyping

We generated a pileup using samtools mpileup (version 1.3), after removing reads with mapping quality lower than 30, bases with base quality lower than 30, and deactivating per-Base Alignment Quality (-q30 -Q30 -B). This pileup was then used to create two separate genotype datasets:

a) For analyses using a SNP panel, pileupCaller (sequenceTools version 1.4.0.3, https://github.com/stschiff/sequenceTools.git) was used. Pseudo-haploid calls were

made for each individual using a random draw approach. To protect our analyses from the effects of DNA deamination damage, we restricted genotyping to transversions only, and excluded transitions from all analyses.

b) For analyses involving rare alleles, an alternative genotyping strategy was used. On positions covered by at least 3 reads, we randomly sampled 3 reads (without replacement) and then carried out a majority call on the sampled reads to produce haploid genotypes. This approach (which is implemented in pileupCaller) ensures all rare alleles are supported by at least two reads, thus reducing error rates, and ameliorates the effects of reference bias in majority calling from high-coverage data(Flegontov et al., 2019).

Sex determination and uniparental markers

Sex determination

The genetic sex of each ancient individual was calculated based on the coverage ratio of sex chromosomes to autosomes. These ratios as well as their associated error bars were calculated using SexDetERRmine (version 1.0, https://github.com/TCLamnidis/Sex.DetERRmine) (Lamnidis et al., 2018), and were calculated using only reads overlapping with variable positions in the 1240k capture panel.

Y chromosome haplotyping

Y chromosome variants were called with ANGSD-0.916 (Korneliussen et al. 2014) using haploid calling function --doHaploCall from 8.8Mb regions uniquely mappable with short read sequencing (Karmin et al., 2015). Haplogroup assignments were made on the basis of *in silico* genotyping of previously published haplogroup-informative markers (David Poznik, 2016; Karmin et al., 2015; Scheib et al., 2018).

Contamination estimation

Autosomal contamination estimate

For male individuals, we used the rate of heterozygosity on the X chromosome to estimate autosomal contamination, using the ANGSD software package (version 0.910) (Rasmussen et al., 2011).

Mitochondrial contamination estimate

We used ContamMix (version 1.0-10)(Fu et al., 2013) to estimate mitochondrial contamination for each ancient individual in our dataset. Reads mapping to the mitochondrial sequence of the reference were extracted using samtools (version 1.3). A consensus sequence was generated from these reads for all positions with coverage higher than 5, using Geneious (version 10.0.9, http://www.geneious.com)(Kearse et al., 2012). Extracted mitochondrial sequences were mapped to their respective consensus sequence using bwa aln/samse (version 0.7.12-r1039). This read alignment, as well a multiple alignment of the consensus sequence to 311 reference

mitochondrial genomes(Fu et al., 2013) were used to run ContamMix, with the --trimbases parameter set to trim 10 bp on either side of the fragments to ameliorate the effects of ancient DNA damage.

Principal component analysis

We used smartpca (version #16000)(Patterson et al., 2006) (https://github.com/DReichLab/EIG) to carry out Principal Component Analysis (PCA), using the "lsqproject: YES" and "shrinkmode: YES" parameters.
For the Worldwide PCA (Figure 2A), the following populations were used to construct principal components: Adygei, Albanian, Aleut, Altaian, Altais, Ami, Armenian, Atayal, Australian, Balochi, Bashkirs, Basque, BedouinB, Bengali, Bergamo, Bougainville, Brahmin, Brahui, Bulgarian, Burmese, Burusho, Buryats, Cambodian, Chane, Chechen, Chipewyan, Chukchi, Cree, Crete, Czech, Dai, Daur, Druze, Dungan, Dusun, English, Eskimo_Chaplin, Eskimo_Naukan, Eskimo_Sireniki, Estonian, Even, Finnish, French, Georgian, Greek, Han, Hawaiian, Hazara, Hazaras, Hezhen, Hungarian, Icelandic, Iranian, Iraqi_Jew, IrtyshBarabinskTatars1, Irula, Itelman, Japanese, Jordanian, Kalash, Kalmyks, Kapu, Karakalpaks, Karitiana, Kashmiri_Pandit, Kazakhs, Kazkahs, Kharia, Khonda_Dora, Kinh, Korean, Kurumba, Kusunda, Kyrgyz, Lahu, Lezgin, Madiga, Makrani, Mala, Mansi, Maori, Mayan, Mexico_Zapotec, Miao, Mixe, Mixtec, Mongola, Nahua, Naxi, Onge, Orcadian, Oroqen, Palestinian, Papuan, Pathan, Piapoco, Pima, Polish, Punjabi, Quechua, Relli, Russia_Abkhasian, Russian, Russia_North_Ossetian, Saami, Samaritan, Sardinian, She, Sherpa, Sindhi, Spanish, Surui, Tajik, Tajiks, Teleuts, Thai, Tibetan, Tlingit, TomskTatars, Tubalar, Tu, Tujia, Turkish, Turkmens, Tuscan, Ulchi, Uyghurs, Uygur, Uzbeks, VolgaTatars, Xibo, Yadava, Yakut, Yemenite_Jew, Yi.

ADMIXTURE analysis

We used ADMIXTURE(Alexander et al., 2009) (version 1.3.0). The program was run on genotype data that was first filtered to exclude variants with a minor allele frequency equal to or below 0.01, and then LD pruned (window size 200, step size 5, and R^2 of 0.5) with plink (v1.90b3.29). Five replicates were run for each K value between 2 and 13, using the --cv option.
The populations used in the analysis were: Adygei, Albanian, Aleut, Altaian, Altais, Ami, Armenian, Atayal, Australian, Aymara_Ventilla, Balochi, BantuHerero, BantuKenya, BantuTswana, Bashkirs, Basque, BedouinB, Bengali, Bergamo, Biaka, Bougainville, Brahmin, Brahui, Bulgarian, Burmese, Burusho, Buryats, Cambodian, Chane, Chechen, Chipewyan, Chukchi, Cree, Crete, Czech, Dai, Daur, Dinka, Druze, Dungan, Dusun, English, Esan, Eskimo_Chaplin, Eskimo_Naukan, Eskimo_Sireniki, Estonian, Even, Finnish, French, Fuegian, Gambian, Georgian, Greek, Han, Hawaiian, Hazara, Hazaras, Hezhen, Hungarian, Icelandic, Igbo, Igorot, Iranian, Iraqi_Jew, IrtyshBarabinskTatars1, Irula, Itelman, Japanese, Jordanian, Ju_hoan_North, Kalash, Kalmyks, Kapu, Karakalpaks, Karitiana, Kashmiri_Pandit, Kazakhs, Kazkahs, Kharia, Khomani_San, Khonda_Dora, Kinh, Kongo, Korean, Kurumba, Kusunda, Kyrgyz, Lahu, Lemande, Lezgin, Luhya, Luo, Madiga, Makrani, Mala, Mandenka, Mansi, Maori, Masai, Mayan, Mbuti, Mende, Mexico_Zapotec, Miao, Mixe, Mixtec, Mongola, Mozabite, Nahua, Naxi, Onge, Orcadian, Oroqen, Palestinian, Papuan, Pathan, Piapoco, Pima, Polish, Punjabi, Quechua, Relli, Russia_Abkhasian, Russia_North_Ossetian, Russian, Saami, Saharawi, Samaritan, Sardinian, She, Sherpa, Sindhi, Somali, Spanish, Surui, Tajik, Tajiks, Teleuts, Thai,

Tibetan, Tlingit, TomskTatars, Tu, Tubalar, Tujia, Turkish, Turkmens, Tuscan, Ulchi, Uyghurs, Uygur, Uzbeks, VolgaTatars, Xibo, Yadava, Yakut, Yemenite_Jew, Yi, Yoruba.

F statistics

We used the Admixtools software package (https://github.com/DReichLab/AdmixTools)(Patterson et al., 2012) for all F statistics related analyses.
qp3Pop (version 435) was used to calculate outgroup F3 statistics, and qpWave (version 410) was used for population continuity analysis.

Rarecoal

We used rarecoal (version 1.5.0.2, https://github.com/stschiff/rarecoal) to model the relationship of ancient and modern South American populations. To generate the tree of modern populations, only variants that were encountered between 2 and 4 times among these six (super-)populations were considered. After the tree was finalised, and rough parameter estimates were provided by the maximum likelihood approach (rarecoal maxl), the parameters were optimised using an MCMC approach (rarecoal mcmc). For the final MCMC parameter optimisation, rare variants encountered between 2 and 10 times among the (super-)populations were considered.

References

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Busing, F. M. T. A., Meijer, E., & Leeden, R. V. D. (1999). Delete-m Jackknife for Unequal m. *Statistics and Computing*, *9*(1), 3–8. https://doi.org/10.1023/A:1008800423698

David Poznik, G. (2016). Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. In *bioRxiv* (p. 088716). https://doi.org/10.1101/088716

de la Fuente, C., Ávila-Arcos, M. C., Galimany, J., Carpenter, M. L., Homburger, J. R., Blanco, A., Contreras, P., Cruz Dávalos, D., Reyes, O., San Roman, M., Moreno-Estrada, A., Campos, P. F., Eng, C., Huntsman, S., Burchard, E. G., Malaspinas, A.-S., Bustamante, C. D., Willerslev, E., Llop, E., … Moraga, M. (2018). Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(17), E4006–E4012. https://doi.org/10.1073/pnas.1715688115

Flegontov, P., Altınışık, N. E., Changmai, P., Rohland, N., Mallick, S., Adamski, N., Bolnick, D. A., Broomandkhoshbacht, N., Candilio, F., Culleton, B. J., Flegontova, O., Friesen, T. M., Jeong, C., Harper, T. K., Keating, D., Kennett, D. J., Kim, A. M., Lamnidis, T. C., Lawson, A. M., … Schiffels, S. (2019). Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature*, *570*(7760), 236–240. https://doi.org/10.1038/s41586-019-1251-y

Flegontov, P., Changmai, P., Zidkova, A., Logacheva, M. D., Altınışık, N. E., Flegontova, O., Gelfand, M. S., Gerasimov, E. S., Khrameeva, E. E., Konovalova, O. P., Neretina, T., Nikolsky, Y. V., Starostin,

G., Stepanova, V. V., Travinsky, I. V., Tříska, M., Tříska, P., & Tatarinova, T. V. (2016). Genomic study of the Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Scientific Reports*, *6*, 20768. https://doi.org/10.1038/srep20768

Fu, Q., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., Ronchitelli, A. M., Martini, F., Cremonesi, R. G., Svoboda, J., Bauer, P., Caramelli, D., Castellano, S., Reich, D., Pääbo, S., & Krause, J. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology: CB*, *23*(7), 553–559. https://doi.org/10.1016/j.cub.2013.02.044

Jeong, C., Balanovsky, O., Lukianova, E., Kahbatkyzy, N., Flegontov, P., Zaporozhchenko, V., Immel, A., Wang, C.-C., Ixan, O., Khussainova, E., Bekmanov, B., Zaibert, V., Lavryashina, M., Pocheshkhova, E., Yusupov, Y., Agdzhoyan, A., Koshel, S., Bukin, A., Nymadawa, P., … Krause, J. (2019). The genetic history of admixture across inner Eurasia. *Nature Ecology & Evolution*, *3*(6), 966–976. https://doi.org/10.1038/s41559-019-0878-2

Karmin, M., Saag, L., Vicente, M., Wilson Sayres, M. A., Järve, M., Talas, U. G., Rootsi, S., Ilumäe, A.-M., Mägi, R., Mitt, M., Pagani, L., Puurand, T., Faltyskova, Z., Clemente, F., Cardona, A., Metspalu, E., Sahakyan, H., Yunusbayev, B., Hudjashov, G., … Kivisild, T. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Research*, *25*(4), 459–466. https://doi.org/10.1101/gr.186684.114

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* , *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Lamnidis, T. C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., Khartanovich, V., Balanovsky, O., Ongyerth, M., Weihmann, A., Sajantila, A., Kelso, J., Pääbo, S., Onkamo, P., Haak, W., Krause, J., & Schiffels, S. (2018). Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nature Communications*, *9*(1), 5018. https://doi.org/10.1038/s41467-018-07483-5

Lindo, J., Haas, R., Hofman, C., Apata, M., Moraga, M., Verdugo, R. A., Watson, J. T., Viviano Llave, C., Witonsky, D., Beall, C., Warinner, C., Novembre, J., Aldenderfer, M., & Di Rienzo, A. (2018). The genetic prehistory of the Andean highlands 7000 years BP though European contact. *Science Advances*, *4*(11), eaau4921. https://doi.org/10.1126/sciadv.aau4921

Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., Nordenfelt, S., Valdiosera, C., Richards, S. M., Rohrlach, A., Romero, M. I. B., Espinoza, I. F., Cagigao, E. T., Jiménez, L. W., Makowski, K., Reyna, I. S. L., Lory, J. M., Torrez, J. A. B., Rivera, M. A., … Haak, W. (2016). Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances*, *2*(4), e1501385. https://doi.org/10.1126/sciadv.1501385

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., Skoglund, P., Lazaridis, I., Sankararaman, S., Fu, Q., Rohland, N., Renaud, G., Erlich, Y., Willems, T., Gallo, C., … Reich, D. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, *538*(7624), 201–206. https://doi.org/10.1038/nature18964

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E. R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J. L., de Castro, J. M. B., Carbonell, E., … Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*(7583), 499–503. https://doi.org/10.1038/nature16152

Moreno-Mayar, J. V., Potter, B. A., Vinner, L., Steinrücken, M., Rasmussen, S., Terhorst, J., Kamm, J. A., Albrechtsen, A., Malaspinas, A.-S., Sikora, M., Reuther, J. D., Irish, J. D., Malhi, R. S., Orlando, L., Song, Y. S., Nielsen, R., Meltzer, D. J., & Willerslev, E. (2018). Terminal Pleistocene Alaskan

genome reveals first founding population of Native Americans. *Nature*. https://doi.org/10.1038/nature25173

Moreno-Mayar, J. V., Vinner, L., de Barros Damgaard, P., de la Fuente, C., Chan, J., Spence, J. P., Allentoft, M. E., Vimala, T., Racimo, F., Pinotti, T., Rasmussen, S., Margaryan, A., Iraeta Orbegozo, M., Mylopotamitaki, D., Wooller, M., Bataille, C., Becerra-Valdivia, L., Chivall, D., Comeskey, D., … Willerslev, E. (2018). Early human dispersals within the Americas. *Science*. https://doi.org/10.1126/science.aav2621

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), e190. https://doi.org/10.1371/journal.pgen.0020190

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., Mendoza, M. L. Z., Beaudoin, A. B., Zutter, C., Larsen, N. K., Potter, B. A., Nielsen, R., Rainville, R. A., Orlando, L., Meltzer, D. J., Kjær, K. H., & Willerslev, E. (2016). Postglacial viability and colonization in North America's ice-free corridor. *Nature*, *537*(7618), 45–49. https://doi.org/10.1038/nature19085

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., & Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. *Genome Biology*, *17*, 60. https://doi.org/10.1186/s13059-016-0918-z

Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., Lamnidis, T. C., Rohland, N., Nägele, K., Adamski, N., Bertolini, E., Broomandkhoshbacht, N., Cooper, A., Culleton, B. J., Ferraz, T., Ferry, M., Furtwängler, A., Haak, W., Harkins, K., Harper, T. K., … Reich, D. (2018). Reconstructing the Deep Population History of Central and South America. *Cell*, *175*(5), 1185–1197.e22. https://doi.org/10.1016/j.cell.2018.10.027

Potter, B. A., Baichtal, J. F., Beaudoin, A. B., Fehren-Schmitz, L., Haynes, C. V., Holliday, V. T., Holmes, C. E., Ives, J. W., Kelly, R. L., Llamas, B., Malhi, R. S., Miller, D. S., Reich, D., Reuther, J. D., Schiffels, S., & Surovell, T. A. (2018). Current evidence allows multiple models for the peopling of the Americas. *Science Advances*, *4*(8), eaat5473. https://doi.org/10.1126/sciadv.aat5473

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T. W., Jr, Orlando, L., Metspalu, E., Karmin, M., Tambets, K., Rootsi, S., Mägi, R., Campos, P. F., Balanovska, E., Balanovsky, O., Khusnutdinova, E., Litvinov, S., … Willerslev, E. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, *505*(7481), 87–91. https://doi.org/10.1038/nature12736

Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M. C., Malaspinas, A.-S., Eriksson, A., Moltke, I., Metspalu, M., Homburger, J. R., Wall, J., Cornejo, O. E., Moreno-Mayar, J. V., Korneliussen, T. S., Pierre, T., … Willerslev, E. (2015). POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*, *349*(6250), aab3884. https://doi.org/10.1126/science.aab3884

Rasmussen, M., Anzick, S. L., Waters, M. R., Skoglund, P., DeGiorgio, M., Stafford, T. W., Jr, Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S. M., Poznik, G. D., Gudmundsdottir, V., Yadav, R., Malaspinas, A.-S., White, S. S., 5th, Allentoft, M. E., Cornejo, O. E., Tambets, K., Eriksson, A., … Willerslev, E. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature*, *506*(7487), 225–229. https://doi.org/10.1038/nature13025

Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K. E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., Kivisild, T., Zhai, W., Eriksson, A., Manica, A., Orlando, L., De La Vega, F. M., Tridico, S., Metspalu, E., Nielsen, K., … Willerslev, E. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science*, *334*(6052), 94–98. https://doi.org/10.1126/science.1211177

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M. V., Rojas, W.,

Duque, C., Mesa, N., García, L. F., Triana, O., Blair, S., Maestre, A., Dib, J. C., Bravi, C. M., Bailliet, G., Corach, D., Hünemeier, T., … Ruiz-Linares, A. (2012). Reconstructing Native American population history. *Nature*, *488*(7411), 370–374. https://doi.org/10.1038/nature11258

Scheib, C. L., Li, H., Desai, T., Link, V., Kendall, C., Dewar, G., Griffith, P. W., Mörseburg, A., Johnson, J. R., Potter, A., Kerr, S. L., Endicott, P., Lindo, J., Haber, M., Xue, Y., Tyler-Smith, C., Sandhu, M. S., Lorenz, J. G., Randall, T. D., … Kivisild, T. (2018). Ancient human parallel lineages within North America contributed to a coastal expansion. *Science*, *360*(6392), 1024–1027. https://doi.org/10.1126/science.aar6851

Schiffels, S., Haak, W., Paajanen, P., Llamas, B., Popescu, E., Loe, L., Clarke, R., Lyons, A., Mortimer, R., Sayer, D., Tyler-Smith, C., Cooper, A., & Durbin, R. (2016). Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nature Communications*, *7*, 10408. https://doi.org/10.1038/ncomms10408

Skoglund, P., Mallick, S., Bortolini, M. C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M. L., Salzano, F. M., Patterson, N., & Reich, D. (2015). Genetic evidence for two founding populations of the Americas. *Nature*, *525*(7567), 104–108. https://doi.org/10.1038/nature14895

6. Manuscript C

Article

Current Biology

Origin and Health Status of First-Generation Africans from Early Colonial Mexico

Highlights

- Genomic and isotopes data suggest an African origin for the three individuals
- One ~14X *Treponema pallidum* sub. *pertenue* genome was recovered
- One ~1,500X hepatitis B virus genome was recovered
- Both pathogen genomes cluster together with present day pathogens from Africa

Authors

Rodrigo Barquera, Thiseas C. Lamnidis, Aditya Kumar Lankapalli, ..., Denise Kühnert, Lourdes Márquez-Morfín, Johannes Krause

Correspondence

kuehnert@shh.mpg.de (D.K.), rlmorfin@gmail.com (L.M.-M.), krause@shh.mpg.de (J.K.)

In Brief

Barquera et al. analyze individuals from a colonial period burial from Mexico. Using an interdisciplinary approach, they reconstruct the genetic ancestry, origins, and health status of three enslaved Africans. Genomes of pathogens recovered from them provide insight on infectious diseases brought to the Americas by the transatlantic slave trade.

Barquera et al., 2020, Current Biology *30*, 2078–2091
June 8, 2020
https://doi.org/10.1016/j.cub.2020.04.002

CellPress

Current Biology

Article

Origin and Health Status of First-Generation Africans from Early Colonial Mexico

Rodrigo Barquera,[1,2,13] Thiseas C. Lamnidis,[1,13] Aditya Kumar Lankapalli,[1,13] Arthur Kocher,[3,13] Diana I. Hernández-Zaragoza,[2,4,13] Elizabeth A. Nelson,[1,5] Adriana C. Zamora-Herrera,[6] Patxi Ramallo,[7,8] Natalia Bernal-Felipe,[9] Alexander Immel,[1,10] Kirsten Bos,[1] Víctor Acuña-Alonzo,[2] Chiara Barbieri,[11,12] Patrick Roberts,[7] Alexander Herbig,[1] Denise Kühnert,[3,*] Lourdes Márquez-Morfín,[6,*] and Johannes Krause[1,14,*]

[1]Department of Archaeogenetics (DAG), Max-Planck Institute for the Science of Human History (MPI-SHH), Kahlaische Str. 10, 07745 Jena, Germany

[2]Molecular Genetics Laboratory, Escuela Nacional de Antropología e Historia (ENAH), Periférico Sur y Zapote s/n. Col. Isidro Fabela, Tlalpan, 14030 Mexico City, Mexico

[3]Transmission, Infection, Diversification & Evolution Group (TIDE), Max-Planck Institute for the Science of Human History (MPI-SHH), Kahlaische Str. 10, 07745 Jena, Germany

[4]Immunogenetics Unit, Técnicas Genéticas Aplicadas a la Clínica (TGAC), Calz. del Hueso 714, Coapa, Los Sauces, Coyoacán, 04940 Mexico City, CDMX, Mexico

[5]Institute for the Archaeological Sciences, University of Tübingen, Geschwister-Scholl-Platz, 72074 Tübingen, Germany

[6]Osteology Laboratory, Post Graduate Studies Division, Escuela Nacional de Antropología e Historia (ENAH), Periférico Sur y Zapote s/n. Col. Isidro Fabela, Tlalpan, 14030 Mexico City, Mexico

[7]Department of Archaeology (DA), Max-Planck Institute for the Science of Human History (MPI-SHH), Kahlaische Str. 10, 07745 Jena, Germany

[8]Faculty of Medicine and Nursing, University of the Basque Country (UPV/EHU), Arriola Pasealekua, 2, 20018 Donostia, Gipuzkoa, Spain

[9]Escuela Nacional de Antropología e Historia (ENAH), Periférico Sur y Zapote s/n. Col. Isidro Fabela, Tlalpan, 14030 Mexico City, Mexico

[10]Institute of Clinical Molecular Biology, Kiel University, Rosalind-Franklin-Straße 12, 24105 Kiel, Germany

[11]Department of Linguistic and Cultural Evolution (DLCE), Max-Planck Institute for the Science of Human History (MPI-SHH), Kahlaische Str. 10, 07745 Jena, Germany

[12]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

[13]These authors contributed equally

[14]Lead Contact

*Correspondence: kuehnert@shh.mpg.de (D.K.), rlmorfin@gmail.com (L.M.-M.), krause@shh.mpg.de (J.K.)

https://doi.org/10.1016/j.cub.2020.04.002

SUMMARY

The forced relocation of several thousand Africans during Mexico's historic period has so far been documented mostly through archival sources, which provide only sparse detail on their origins and lived experience. Here, we employ a bioarchaeological approach to explore the life history of three 16th century Africans from a mass burial at the San José de los Naturales Royal Hospital in Mexico City. Our approach draws together ancient genomic data, osteological analysis, strontium isotope data from tooth enamel, $\delta^{13}C$ and $\delta^{15}N$ isotope data from dentine, and ethnohistorical information to reveal unprecedented detail on their origins and health. Analyses of skeletal features, radiogenic isotopes, and genetic data from uniparental, genome-wide, and human leukocyte antigen (HLA) markers are consistent with a Sub-Saharan African origin for all three individuals. Complete genomes of *Treponema pallidum* sub. *pertenue* (causative agent of yaws) and hepatitis B virus (HBV) recovered from these individuals provide insight into their health as related to infectious disease. Phylogenetic analysis of both pathogens reveals their close relationship to strains circulating in current West African populations, lending support to their origins in this region. The further relationship between the treponemal genome retrieved and a treponemal genome previously typed in an individual from Colonial Mexico highlights the role of the transatlantic slave trade in the introduction and dissemination of pathogens into the New World. Putting together all lines of evidence, we were able to create a biological portrait of three individuals whose life stories have long been silenced by disreputable historical events.

INTRODUCTION

Almost 500 years ago, in 1518, Charles I of Spain issued an authorization to transport the first African slaves into the Viceroyalty of New Spain (which at its zenith comprised the entirety of present day Mexico, the Caribbean, parts of the United States and Canada, and all Central America except for Panama) [1]. 5 centuries later, the ancestry of those hundreds of thousands of forcefully abducted people has formed an integral part of the genetic heritage carried by a large number of people in Mexico and a highly visible part of its national cultural heritage. Slavery was the primary mechanism of immigration of Africans to Mexico

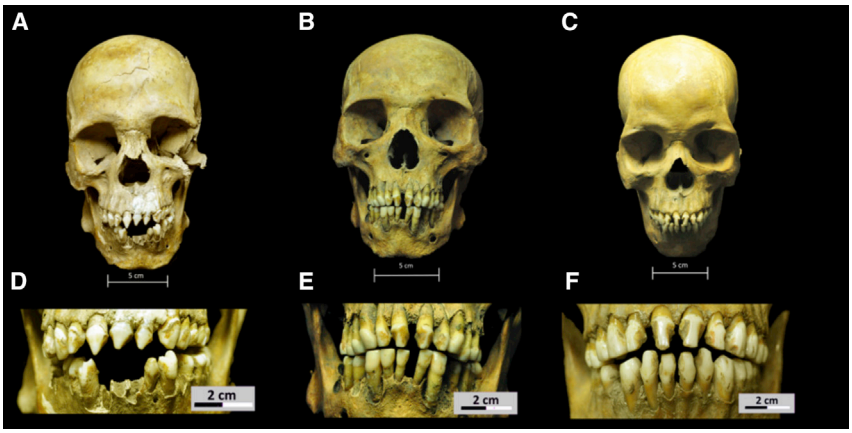


Figure 1. Skulls and Dental Decoration Patterns for the Three African Individuals from the San José de los Naturales Royal Hospital

(A) Skull from individual 150 (SJN001).
(B) Skull from individual 214 (SJN002).
(C) Skull from individual 296 (SJN003).
(D) Close up of dental modification patterns for individual 150 (SJN001).
(E) Close up of dental modification patterns for individual 214 (SJN002).
(F) Close up of dental modification patterns for individual 296 (SJN003).
The skeletons of these individuals are part of the collection of San José de los Naturales, guarded at the Osteology Laboratory of the Post Graduate Studies Division at the National School of Anthropology and History (ENAH), Mexico City, Mexico. Photographs: R. Barquera and N. Bernal.

during the first half of the colonial period, which constitutes one of the most significant forced oppressive migrations in human history [2]. The increased demand for enslaved manual laborers, together with the establishment of the first European settlements in what is euro-centrically referred to as the "New World," caused the growth and consolidation of the transatlantic slave trade. Despite uncertainties around absolute numbers, an estimated 10.6 to 19.4 million Africans had been forcibly deported from their homelands until slavery was finally abolished in most parts of the Americas in the 1860s. To add to these staggering figures, high mortality rates during their voyages meant that only 9.6 to 15.5 million actually arrived in the Americas [3–6]. These data include only the entries through the official ports, which ignores individuals that were smuggled, for whom there is no reliable record. In the case of the Viceroyalty of New Spain, between 130,000 and 150,000 Africans had arrived by 1779, the year in which the importation of slaves into New Spain was banned [7, 8]. Of these, approximately 70,000 entered between the years 1600 and 1640 [3]. The sudden increase in demand was in part due to a reduction in the indigenous labor force that resulted from both casualties in the many conflicts during the European conquest and from diseases (among them, smallpox, measles, and typhoid fever) [9–11] that devastated nearly 90% of the native population [10]. Creoles, Africans, mulattoes, and other African-descended groups were thought to have higher resistance to these diseases compared to Indigenous Americans and Europeans [7, 12–14] making them desirable assets. Further to this, *Las Leyes Nuevas* (The New Laws) of 1542 prohibited the use of Native American labor as slaves in New Spain [15].

The African genetic diversity has been studied in other regions of the Americas by means of genome-wide analysis of both ancient [16–18] and present-day [19, 20] populations. To better understand the lives and health status of first-generation African slaves in Colonial Mexico, historical, archaeological, isotopic, and genetic evidence needs to be considered. Here, we use a bioarchaeological approach that marries ancient human genetic data with osteological analyses, stable carbon and nitrogen isotope analysis, strontium isotope analysis, ethnohistorical information, as well as a molecular screening for potential pathogens to comment on the origins, health status, and life history of three putatively enslaved Africans from Mexico City from a mass burial at the *Hospital Real de San José de los Naturales* (San José de los Naturales Royal Hospital [SJN]) [10, 21]. Documentary evidence indicates this hospital serviced exclusively *los naturales* (i.e., the indigenous population) [22]; however, different sources and studies point to non-indigenous peoples as being treated and buried there [23–27]. Particular traits, such as culturally associated dental modification patterns (Figure 1; see also [6, 28]), suggested a possible African origin of three individuals. Indeed, our results confirm and characterize their African ancestry. We got important insights into these individuals' health status and retrieved genomic material of two pathogens phylogenetically linked to strains from Sub-Saharan Africa. Based on isotopic evidence, we can suggest a non-Mexican origin for these individuals, as they are presumably first-generation enslaved Africans who died very early during the colonial period in Mexico.

RESULTS

Samples were processed according to protocols designed for ancient DNA work in dedicated facilities. We were able to extract DNA from tooth dentine powder for all three individuals, which was used for building genomic libraries and producing ~5 million reads for each of them. Pre-processed sequencing data were used for pathogen screening, and PCR reamplified, reconditioned libraries were enriched for human-specific genomic regions or genomes of pathogens. After sequencing the enriched genomic libraries, quality control (QC), population genetics, and pathogen genome reconstruction analyses were carried out.

Ancient DNA Authentication

We observed short average fragment length (~55 bp) and the characteristic increase in cytosine (C) to thymine (T) transitions toward the termini of the DNA fragments that support the presence of authentic ancient DNA (aDNA) in the dataset. Pre-processed sequences were mapped to the human genome assembly GRCh37 (hg19) [29]. For our uracil DNA glycosylase (UDG)-treated libraries, the rate of deamination damage in the first base on the 5′ and 3′ termini were ~7% and 6%, respectively. After enriching for 1,237,207 single-nucleotide polymorphisms (SNPs), we decreased the number of SNPs to place our coverage in the range of that we generated for the ~600,000 SNP positions present in the Human Origins dataset [30], where ~400,000 SNPs for the three enriched libraries

were obtained. Coverage comparisons of X and Y chromosomes SNPs [31] from the 1,240K capture set assigned all three individuals as genetically male. To estimate levels of nuclear contamination in our samples, we used *ANGSD* to measure the rate of heterozygosity of polymorphic sites on the X chromosome [32]. Further to this, a reconstructed mtDNA consensus sequence was used to estimate mtDNA contamination levels using *schmutzi* [33]. Both methods yielded comparable contamination estimates below 2% for all three individuals (Table 1). The pairwise mismatch rates do not support genetic kinship among the three individuals (Table 2).

Characterization of African Ancestry

The carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) measurements of dentine (Table S2) provided values between $-19.9‰$ and $-18.8‰$ ($\delta^{13}C$) and from $13.2‰$ to $10.8‰$ ($\delta^{15}N$). Broadly speaking, those values show that the dietary protein of the three individuals was composed mainly of terrestrial proteins but with a significant portion of C_4 plants and/or marine proteins, consistent with documented diets either in the arid interior or the western coast of West Africa [34]. Unfortunately, it is difficult to distinguish the two kinds of diet due to an absence of associated faunal remains with the individuals that would provide a local ecological "baseline." A comparison between individuals shows higher nitrogen values ($\delta^{15}N = 13.2‰$) for individual SJN001 relative to the other two individuals ($\delta^{15}N = 10.8‰$ and $10.5‰$ for SJN002 and SJN003, respectively). Without any additional contextual information or isotopic data from fauna, it remains challenging to determine the exact nature of their individual diets. Distinguishing these sources has been a consistent problem in African archaeology, where significant C_4 ecosystems coexist in proximity to the coast [35, 36].

Osteological analyses of the three individuals reveal evidence suggesting a life experience of conflict and hardship. Individual *ML8 SL 150* (SJN001) was found with five buck shots and two healing needles (used in traditional medicine) in the thoracic cavity, as well as gunshot wounds. Both SJN001 and SJN003 (*ML8 SLU9B 296*) presented porotic hyperostosis and *cribra orbitalia*, two pathological changes associated with a skeletal response to nutritionally inadequate diets, anemia, parasitic infectious diseases, and blood loss [37–40]. Individual *ML8 San José 214* (SJN002) displayed several skeletal changes associated with intense labor and heavy manual activity, including enthesopathies on the clavicle and scapulae as well as osteophytic lipping on the joint surfaces with some additional joint contour deformation at the sternoclavicular joint of the clavicle. Additionally, he suffered from a poorly aligned complete fracture in the right fibula and tibia, resulting in associated joint changes of the knee, including osteochondritis dissecans of the distal femoral surface with joint contour deformation and associated osteophytic lipping of the articular surface margin. Furthermore, this individual displayed osteoarthrosis of the lumbar vertebrae in addition to signs of deficient oral health and cut marks on the frontal bone.

Skeletal analysis of cranial nonmetric traits (please refer to the STAR Methods section for further information) showed a strong African affinity for all three individuals (macromorphoscopic characteristics: 92.23% for SJN001; 95.24% for SJN002; 91.95% for SJN003). The dental modification patterns (i.e., decoration of the tooth by mechanical modification) are consistent with cultural practices previously reported for African enslaved individuals from different contexts [41, 42] but also observed in groups living in Western Africa today, most notably in the coastal area and nearby regions of equatorial Guinea, Gabon, Cameroon, the Republic of Congo, and the Democratic Republic of the Congo [43]. Uniparental haplogroups and human leukocyte antigen (HLA) haplotypes are representative of Sub-Saharan populations (Table 1). The mitochondrial lineages of the SJN individuals were compared to a set of published mitochondrial genomes from the same three haplogroups (Figure S1). The three mtDNA haplogroups found show their highest frequencies in Western and Central Africa [44], with haplogroup L1b (present in SJN001) concentrated in Western-Central Africa, particularly along the coastal areas [44, 45]; L3d (present in SJN002) being prevalent in western Sub-Saharan Africa [44, 46]; and L3e1 (present in SJN003) being prevalent in the central areas of the continent [44, 47, 48]. Regarding the paternal lineages, all three individuals carried the Y chromosome E3b1a (E-M2) lineage (but not the same haplotypes; Table 1), which is highly prevalent in modern Sub-Saharan populations (particularly in Western Africa) and equatorial Africans and is also the most common lineage among African Americans [49–53]. In order to get a glimpse at the immunogenetic diversity present in these individuals, we carried out immune capture to obtain information on the HLA allelic variation. HLA diversity is helpful in determining the susceptibility and resistance to specific diseases, both infectious and autoimmune. HLA typing poses technical difficulties, and it can be challenging to call HLA alleles from ancient DNA data. For that reason, we enriched the HLA region with an in-solution capture approach (please refer to the STAR Methods section for further detail), yielding 10×–100× coverage for the HLA region. Allele assignation was carried out using Optitype [54], a software specifically developed to deal with short reads coming from ancient DNA data. The HLA haplotypes described here have so far been reported only in African or African-descent populations, like Mozambique, Kenya, South Africa, and African Americans [55–59]. In the case of both HLA haplotypes of SJN001 and haplotype HLA-A*74:01∼B*49:01∼C*07:01∼DRB1*04:05∼DRB4*01:01∼DQA1*03:02∼DQB1*03:02∼DPA1*02:01∼DPB1*13:01, present in SJN002 (Table 1), none of these haplotypes were yet reported in any African populations, but only in mixed ancestry individuals of at least partial African descent [59, 60]. However, the observed *HLA-A∼B∼C* blocks (corresponding to the HLA class I genomic region, functionally devoted to present endogenous peptides to the immune system) have been reported in Sub-Saharan African populations, such as Luo from Kenya [57] and in Rwandan women [61].

Genomic data from the three individuals were merged to 593,124 autosomal SNPs of the Human Origins (HO) dataset [30] for downstream population genetics analyses. To further assess the genetic relationship of these African individuals to present-day populations from Africa, we carried out principal-component analysis (PCA). We first projected the data from the three individuals on PCs calculated on variation from 371 worldwide populations (Figure S2A) and then on variation from 534 individuals from 51 populations from Northern and Sub-Saharan Africa. We found that the SJN individuals cluster together with Sub-Saharan African populations like Mandenka, Mende,

Table 1. Basic QC and Genetic Markers of the Analyzed African Individuals from the Royal Hospital of San José de los Naturales

Library	Sex	Modeled ^{14}C Dating	Damage 1st Base 3′	Damage 1st Base 5′	Called SNPs on HO (Post-merging)	Contamination Estimates (Post-merging)	
						Schmutzi	ANGSD
SJN001.A0102	M	AD 1453–1626	0.0678	0.0751	427,155	0.02 ± 0.01	0.54% ± 0.002%
SJN001.B0102			0.0309	0.0429			
SJN002.A0102	M	AD 1450–1620	0.0710	0.0803	397,342	0.02 ± 0.01	0.44% ± 0.002%
SJN002.B0102			0.0776	0.0882			
SJN003.A0102	M	AD 1436–1472[a]	0.0474	0.0730	446,940	0.02 ± 0.01	0.28% ± 0.002%
SJN003.B0102			0.0443	0.0698			

	Uniparental Markers		HLA Haplotypes								
Individual	mtDNA Haplogroup	Y-Chr Haplogroup	HLA-A	HLA-B	HLA-C	HLA-DRB1	HLA-DRB3/4/5	HLA-DQA1	HLA-DQB1	HLA-DPA1	HLA-DPB1
SJN001	L1b2a	E1b1a1a1c1b | E-M263.2	A*03:01	B*44:10	C*04:01	DRB1*07:01	DRB4*01:01	DQA1*02:01	DQB1*02:01	DPA1*02:01	DPB1*01:01
			A*66:02	B*44:03	C*03:03	DRB1*16:02	DRB5*02:02	DQA1*01:02	DQB1*05:02	DPA1*02:02	DPB1*01:01
SJN002	L3d1a1a	E1b1a1a1d1 | E-P278.1 E-M4254	A*30:01	B*42:01	C*17:01	DRB1*03:02	DRB3*01:01	DQA1*04:01	DQB1*04:02	DPA1*02:01	DPB1*15:01
			A*74:01	B*49:01	C*07:01	DRB1*04:05	DRB4*01:01	DQA1*03:02	DQB1*03:02	DPA1*02:01	DPB1*13:01
SJN003	L3e1a1a	E1b1a1a1c1a1c (E-CTS8030)	A*30:01	B*42:01	C*17:01	DRB1*07:01	DRB4*01:01	DQA1*03:01	DQB1*02:02	DPA1*01:03	DPB1*04:01
			A*66:02	B*58:01	C*07:01	DRB1*15:03	DRB5*01:01	DQA1*01:02	DQB1*06:02	DPA1*01:03	DPB1*02:01

[a]These years refer to pre-contact dates; however, a possible reservoir effect due to marine diet cannot be ruled out as discussed in the Results section.

Table 2. Pairwise Mismatch Ratio Analysis

Individual 1	Individual 2	SNPs Tested	Mismatched SNPs	Mismatch Ratio
SJN001	SJN002	594,706	160,514	0.2699
SJN001	SJN003	661,252	178,194	0.2695
SJN002	SJN003	613,943	164,424	0.2678

Wambo, Ovambo, and Damara. When PC1 and PC2 are considered (Figure S2B), the three individuals are projected in the Bantu-speaking population cluster. When plotting PC1 versus PC3 (Figure 2A), the three individuals from SJN appear slightly separated from each other. SJN002 and SJN003 practically overlap with each other and are projected in close proximity to the Wambo population of Southern Africa. SJN001 is projected close to the other two individuals from SJN but falls even closer to the Mende population of Western Africa. PC3 separates hunter-gatherer populations like the Mbuti, Biaka, and Hadza from all other African populations. Admixture estimates (lowest cross validation error at $K = 10$) show that the SJN individuals look very similar, though not identical, in terms of genetic ancestry composition (Figure 2C). Predominantly, ancestry components that are shared with Sub-Saharan populations, such as Yoruba (non-Bantu speakers), Mandenka (non-Atlantic Niger-Congo linguistic family), and Bantu speakers from Kenya and South Africa, were found in all three individuals. An additional component shared with Mbuti and/or Biaka was also evident for SJN002 and SJN003. This extra component might affect the projection of these individuals in the PCA, especially when PC3 is considered. No ancestry related to other continental sources (especially Native Americans) is visible in the ADMIXTURE results (Figure 2C). Altogether, these data support a heterogeneous genetic makeup for the SJN individuals within Sub-Saharan Africa, with clear links to Central-Western Africa and possible links with populations who live in Southern Africa today.

Fine-Scale Population Genetics Analyses

We calculate outgroup F_3 statistics of the form f_3 (outgroup; SJN00X, *Y*) to measure the amount of shared genetic drift of each individual from SJN with a panel of worldwide populations, using the genome of the Ust'-Ishim early modern human from Siberia [62] as an outgroup. Ust' Ishim is a 45,000-year-old Siberian individual that is equally related to both East Asians and West Eurasian hunter-gatherers [62]. This individual represents therefore an outgroup population for West African populations. The highest f_3 values for each individual were with human groups speaking Niger-Congo languages (Figure 3). To test to which specific African population the SJN individuals are most closely related to or shared an excess of alleles with any given population, *D*-statistics of the form *D* (chimp, SJN00X; *X*, target) were used, where target was substituted by the population that gave the highest f_3 value for each individual. For all three individuals from SJN, no population was significantly more closely related to them than to the target population. For SJN001 (Figure S3A), Mende were used as target; for SJN002 (Figure S3B), Bantu-speaking Ovambo were used as target; and for SJN003 (Figure S3C), Lemande were used as target. To test for cladality between each of the individuals from SJN and their respective target populations, we computed *D* statistics of the form *D* (chimp, X; target, SJN00X). We find that none of the ancient individuals are truly cladal with any target population (Table S1). The fact that none of the individuals showed specific genetic affinity with any modern African populations can be due to (1) past migration (i.e., during the colonial period) and/or reallocation of populations, (2) not enough representation of the groups they are more closely related to in the datasets tested, or (3) the populations they are genetically related to do not exist any longer. Interestingly, a recent manuscript [18] described the recovery and characterization of genomic DNA from a tobacco pipe belonging to an African woman that was found to be genetically closely related to present-day Mende from Sierra Leone, which further support Mende-related populations as a source of African genetic ancestry brought into the Americas during the colonial period.

Evidence of the Place of Birth from Strontium Isotope Analysis

Radiocarbon dating suggests that these individuals died shortly after the beginning of the colonial period in Mexico City (Table 1). To further assess the origin of these ancient individuals, we carried out strontium analysis on the molars of each individual, which forms early during childhood and thus is a good proxy for place of birth. For all three individuals, the strontium ratio values are above 0.7089 (SJN001: 0.71078 ± 0.00001; SJN002: 0.72045 ± 0.00001; SJN003: 0.72026 ± 0.00001), which is the highest value recorded for Mexico in the northern part of the Yucatán Peninsula (Tizimin and Dzibilchaltun) [63–65] and even higher than those found for central Mexico (Teotihuacan [0.7049], Xico [0.7045], Chapantongo [0.7052], and Mexico City [0.7062–0.7064]) [63, 66]. In fact, these values are consistent with a West African origin (0.70603–0.74143; Figures 2B and S4; see also [63, 67]).

Paleomicrobiological Findings

The three individuals were buried in a mass grave in the grounds located just outside of the hospital. Their mass grave contained skeletal remains of several individuals disposed in layers, which is consistent with burials made during epidemics, in which dead bodies rapidly outnumber the availability of single graves. We therefore decided to screen these individuals for potential pathogenic agents using a bioinformatic approach [68] to screen and filter reads from the genomic libraries, and this way, we obtained reads mapping to bacteria that could either be part of the oral microbiota (like *Tannerella forsythia*, *Streptococcus mutans*, and *S. gordonii*) or implicated in taphonomic processes (*Clostridium tetani*). Apart from environmental bacteria, 51 reads were mapping specifically to *Treponema pallidum* genomes in individual SJN003, which were marked as potentially positive for treponemal infection. We also found evidence of hepatitis B virus (HBV) DNA in individual SJN001 (library SJN001.B0102), with eight sequencing reads mapping specifically to HBV.

The positive libraries were enriched for HBV (SJN001) and *Treponema pallidum* (SJN003), respectively, using a modified in-solution capture strategy followed by paired-end sequencing (Supplemental Information for further details). After capture, a total of 50,493 (deduplicated) reads mapping to the HBV reference genome were recovered from SJN001. Those exhibited typical patterns of ancient DNA damage, and the complete HBV circular

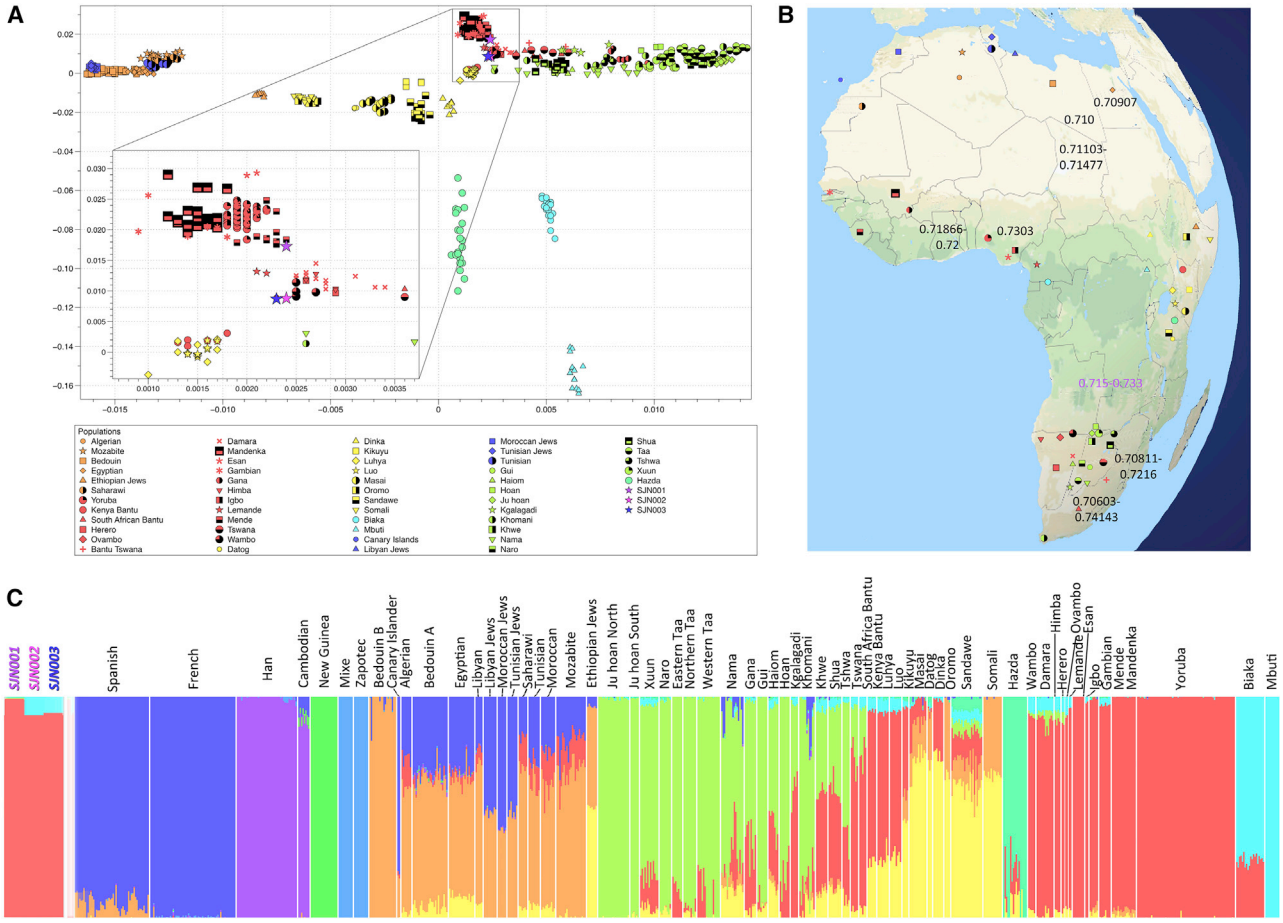


Figure 2. PCA (PC1 versus PC3), Geographic Locations of African Populations Used in the Present Study, and Typical Sr Ratios for Africa

(A) PCA (PC1 versus PC3) showing the genetic relationship of SJN Africans (purple, pink, and violet stars) to Western Africa Niger-Congo linguistic speakers.

(B) Map of Africa showing the places where African populations used for PCA and ADMIXTURE analyses are from. Numbers indicate typical $^{87}Sr/^{86}Sr$ ratios for some regions of Africa (see also Figure S4).

(C) Admixture plot (K = 10) for African populations and potential genetic variation candidate sources due to the demographic history of the Viceroyalty of New Spain (modern Mexico).

genome could be recovered, with a mean coverage of 1,567-fold. These results confirmed that SJN001 was infected with HBV. The ancient HBV genome exhibited a 6-nt insert at the carboxyl end of the core gene, which is characteristic of the HBV A genotype [69]. The phylogenetic analysis confirmed that SJN001 HBV belonged to genotype A (Figure 4) and further showed that it was included in quasi-subgenotype A3 (QS-A3) [70]. Today, QS-A3 is typically found in West Africa. More specifically, SJN001 HBV grouped with strains isolated from Gambia and Guinea.

From SJN003, we reconstructed a *T. pallidum* genome at a mean coverage of 14-fold with 96.66% of the reference genome (Nichols strain) covered at 5-fold average coverage. After aligning the SJN003 genome to previously reported ancient and modern *Treponema pallidum* genomes [71], we produced a maximum likelihood tree (Figure 5) that showed the SJN003 genome to cluster within *Treponema pallidum* subsp. *pertenue* (TPE), the causative agent of yaws. Surprisingly, we found that the SJN003 strain is most closely related to the ancient TPE genome 133 previously isolated from human remains from Colonial Mexico [71]. Both genomes are phylogenetically associated with strains isolated from patients in Ghana.

DISCUSSION

Our combination of genetic analyses, isotope data, ethnohistorical information, and osteobiographies led to the construction of unique life portraits for three individuals from a colonial hospital cemetery in Mexico City. We opted to treat these three individuals independently as opposed to predicating our analysis on the assumption that they represent a singular cultural tradition and ethnic history; this approach was used based on the evidence of forced social intermixing as a means of social control and disintegration among the enslaved population [2, 3, 5, 72]. Osteological assessment, dental modification patterns, and their genetic ancestry are all consistent with a Western or Southern African origin for the three individuals. This finding is further supported by strontium isotope ratios that suggest a non-local origin outside Mexico. The lives of physical hardships and traumas revealed by their osteobiographies, coupled with the fact that most

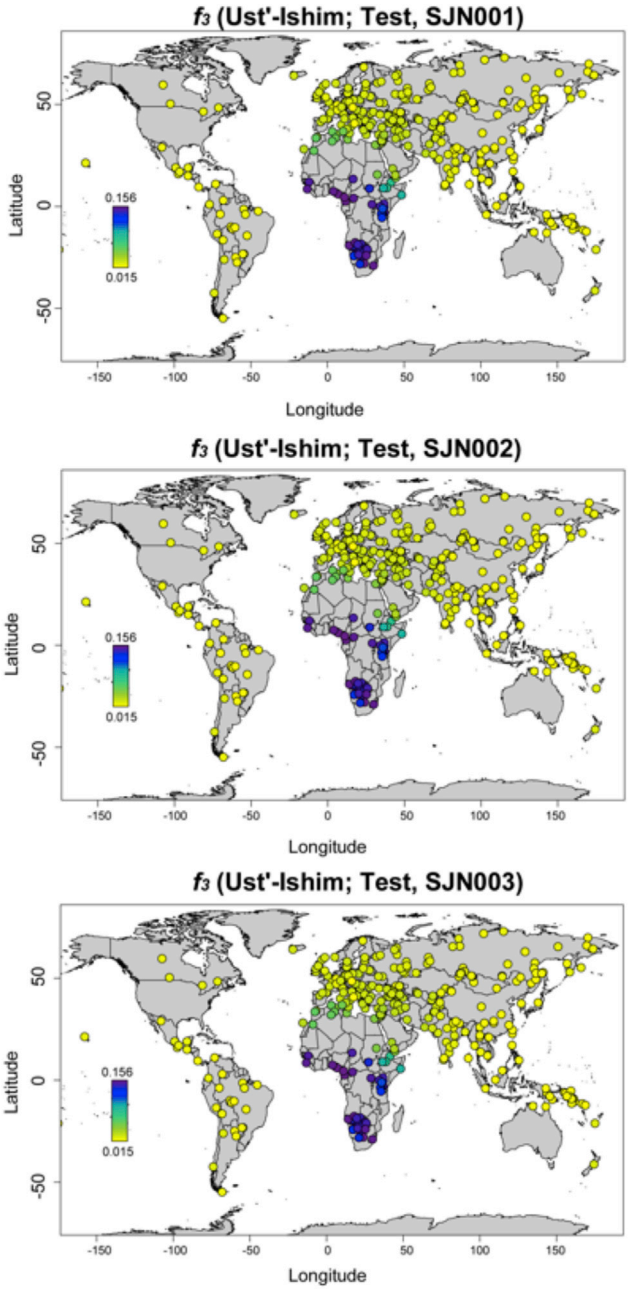


Figure 3. Heatmap of Outgroup f_3 Values for Each Individual from SJN Using a Worldwide Present-Day Population Dataset and Ust'-Ishim as the Outgroup

Higher f_3 values correspond to darker colors.

Sub-Saharan Africans in the Americas from the early colonial period (consistent with SJN individuals' radiocarbon dates) were taken there by force, make the natural conclusion that these individuals were actually enslaved Africans, even in absence of concrete archaeological evidence (such as the individual from the ex-Temple of *Corpus Christi* who was found in shackles) [73].

The genomic analyses suggest that all three individuals show unadmixed African ancestry that is similar to that of Western and Southern African populations. Combining those results with the strontium isotope ratios evidence allows us to conclude that all three individuals were born outside Mexico and likely originate from Western Africa. To the best of our knowledge, they are the earliest genetically identified first-generation Africans in the Americas. mtDNA lineages found in the individuals from SJN (L1 and L3) have been found, although not highly represented, among modern day Mexicans with mixed ancestry [74, 75]. The African HLA haplotypes found in the SJN individuals, or parts of their haplotypes, can be found throughout modern Mexico in frequencies ranging from 0.01% to 2.27% [76]. Alleles HLA-DRB1*07 and HLA-DQB1*02 have been previously found associated with decreased antibody response against HBV and non-responsiveness to HBV vaccination [77, 78]. Interestingly, those two alleles are present in individual SJN001, which was found to be infected with HBV. Those alleles are also present in SJN003 but joined by protective alleles [77, 79], namely HLA-DQB1*06:02 and HLA-DPB1*02:01, which could have aided this individual not to become infected, if he would have been exposed to the virus in the first place. There is currently no reported association between HLA alleles and treponemal infections (apart from HLA-DRB1*14 in Asians) [80]. The fact that three of the six haplotypes were not previously reported in any African population highlights the necessity for exhaustive sampling of HLA haplotypes in African populations. D-statistics identify the closest genetic matches for the three individuals in central West Africa, Western Africa, and Southern African populations (Bantu-speaking groups). Our analyses point to a high genetic diversity in the populations who were the initial sources of the slave trade as suggested by historical records [3, 7, 81]. Furthermore, it is possible that the genetic diversity of living African populations was shaped by substantial migrations through the past four centuries, which displaced human groups and genetic ancestries. The analysis of ancient DNA from early African slave migrants in the Americas therefore provides an alternative opportunity to look at the past genetic makeup of the African continent, where the climatic conditions are not ideal for DNA preservation.

Based on radiocarbon dating, the three individuals lived during the early years of the colonial period in Mexico City (^{14}C range: AD 1436–1626). The osteobiography revealed non-specific markers of physiological stress, evidence of occupational stress and healed severe wounds. They were found in the context of a mass grave (stacked in several layers suggestive of catastrophe deposits made during periods of epidemics) in proximity to a hospital that served only indigenous people [21], from a time period notorious for major epidemics in the region.

Based on our molecular screening, two of the three individuals showed molecular evidence of infectious diseases. HBV is widespread today and poses a significant global health burden [82]. It is currently classified into ten different genotypes that are heterogeneously distributed around the globe [69], and its long history infecting humans dates back at least to the Late Neolithic and Bronze Age [83, 84]. The biogeographic distribution of this virus is tightly linked to human history and migration, given its ease of transmission through contact with infected body fluids and the fact that it can cause a chronic infection [82]. Dispersion of HBV in the past through forced African migration was suggested by the presence of extant African HBV subgenotypes in regions

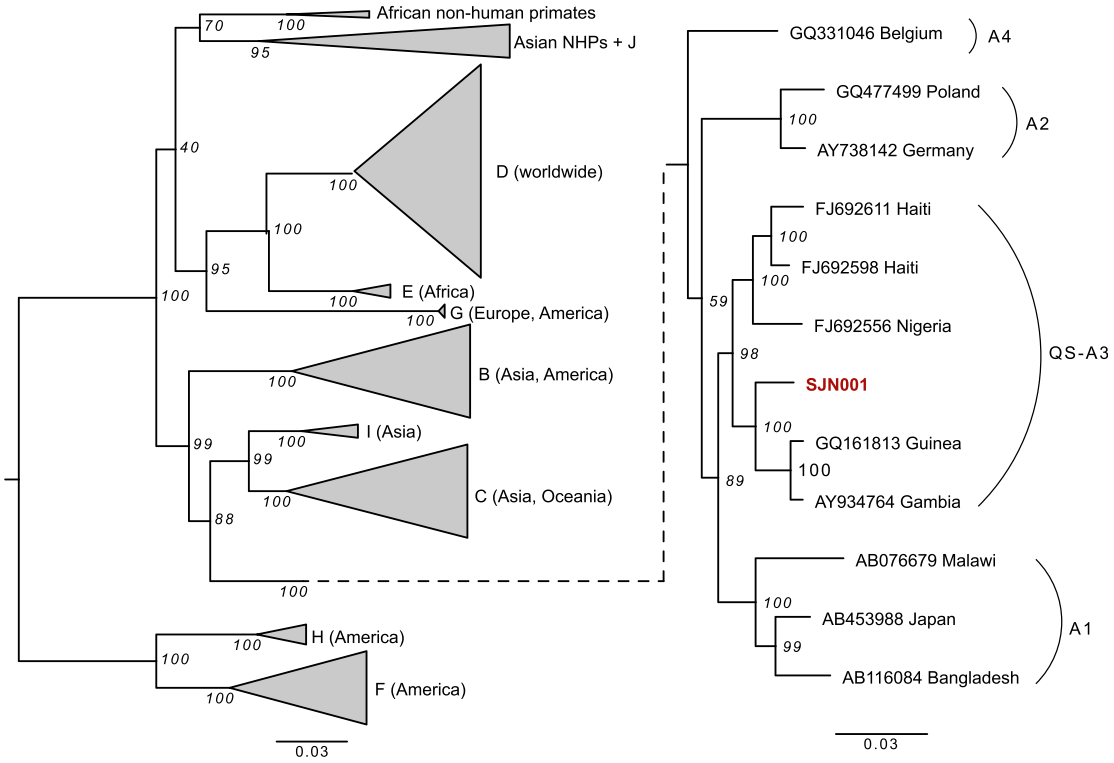


Figure 4. Phylogenetic Tree of Hepatitis B Virus Based on Full Genomes, Including SJN001, Estimated by Maximum Likelihood

Branch lengths represent numbers of substitutions. Bootstrap support values are written at the nodes. In the left panel, inter-genotype relationships are shown and the main continental distribution of each genotype is indicated [69]. The relationships within genotype A and the position of SJN001 are shown in the right panel.

where slavery was practiced [69, 85]. Given that individual SJN001 carried the HBV QS-A3 genotype typically found in West Africa today, coupled with the data that support his origins in this region, provides the first direct evidence of HBV disease movement with a potential dissemination through the transatlantic slave trade. Today, Haiti is the only location outside Africa where the HBV QS-A3 subgenotype is found [69, 85]. Thus, our results reveal that this HBV lineage seems not to have established itself in Mexico.

In contrast, the variety of yaws identified in individual SJN003 shares a common origin with the strain identified in a 17th-century individual of Colonial Mexico City carrying a European mitochondrial haplogroup (H1c+152) [71]. The two strains also cluster with yaws subtypes identified in extant West African groups (although the nodes for the relevant sections of the tree have low statistical support; Figure 5). The osteopathological findings on SJN003 (i.e., cribra orbitalia and hyperostosis and a mild lesion in the frontal bone; please refer to the STAR Methods section for further details) are consistent with chronic infectious diseases and have been reported in suspected cases of individuals affected by treponemal diseases. The polymorphic nature of the lesions left on the bone by treponemal bacteria invites one to pay attention to any potential treponematosis case [71]. These data suggest the establishment of at least one disease of African origin in the local population. As yaws is a highly contagious skin infection that is associated with poor hygiene, it is unsurprising that skeletal signs of treponemal diseases have been reported in other African enslaved individuals from similar historic periods in other contexts in Europe [42]. Our findings add to the discussion of the dynamics of treponemal diseases in the colonial period of Mexico, where they were a prominent yet not completely understood health issue [86, 87]. Our current data, however, shed little light on the reason for their deposition in a presumed epidemic burial ground.

Here, we present the application of a comprehensive set of methods and techniques to display the first in-depth characterization of first-generation African slaves in Colonial Mexico, a historically oppressed group. This was accomplished through reconstruction of the origins and life histories of three putatively enslaved Africans from the beginning of the colonial period in central Mexico. We were able to describe their biological identity and explore elements of their health status by recovering high-quality ancient DNA from both the individuals and their identifiable pathogens. Our work adds important information on the origin, health status, and life histories of the first generation of African slaves that were forcefully relocated and transported to the New World during one of the most horrific cases of callous disregard for human life and violation of human rights during the colonial period. By investigating the origin and disease experience of these individuals through molecular methods and evaluating the skeleton for signs of life experience and cultural affinity, we illuminate, in some measure, the identity, culture, and life of these people whose history has largely been lost. Furthermore, by exploring the African diversity that first came into Mexico and

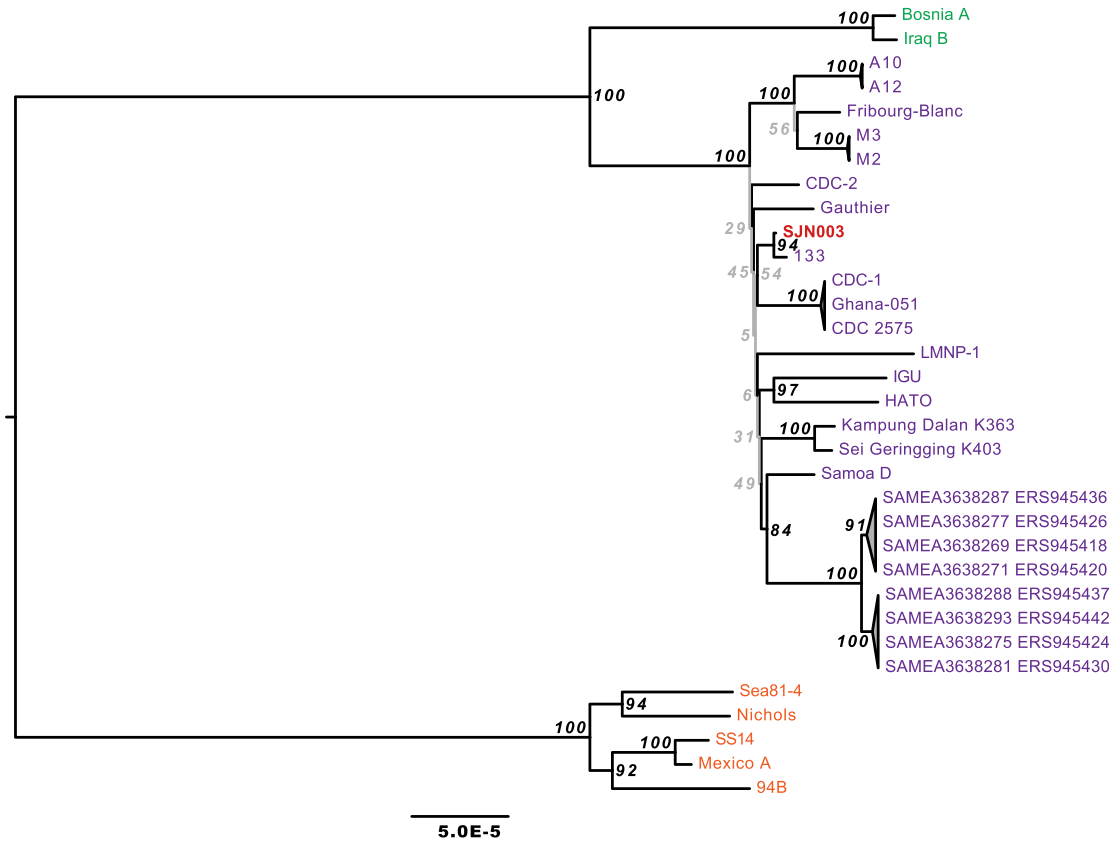


Figure 5. Maximum Likelihood Tree for Ancient Treponemal Genomes from SJN003 (*T. pallidum* sub. *pertenue*), 113 (*T. pallidum* sub. *pertenue*), and 94B SJN003 (*T. pallidum* sub. *pallidum*) and 30 Modern *Treponema pallidum* Genomes

Orange, violet, and green tip labels represent *T. pallidum* subsp. *pallidum* (causative agent of syphilis), *T. pallidum* sub. *pertenue* (causative agent of yaws), and *T. pallidum* sub. *endemicum* (causative agent of bejel) strains, respectively. The scale represents the mean number of substitutions per site for the GTR+GAMMA substitution model. The strain found in individual SJN003 branches with *T. pallidum* sub. *pertenue* strains. Bootstrap support for each branch estimated from 1,000 replicates is provided. The branches with statistical support lower than 80 are presented in gray.

Latin America, we can draw a sharper picture of the biological legacy and cultural roots of modern Latin Americans, explore the extent of this diversity through different disciplines, and also acknowledge how African populations contributed to the biological diversity of mixed ancestry populations of the Americas.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Archaeological sites and sample description
- METHOD DETAILS
  - Osteological assessment
  - Stable isotope analysis, diets, and the reservoir effect
  - Ancient DNA sample processing and quality control
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - QC and data processing for human PopGen
  - Pathogen genomes assemblies

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.04.002.

ACKNOWLEDGMENTS

The authors wish to thank Franziska Aron, Cäcilia Freund, Rita Radzeviciute, Raphaela Stahl, Marta Burri, Antje Wissgott, Guido Brandt, András Szolek, Stephen Clayton, Perla del Carmen Ruíz Albarrán, Laura Curiel Giles, and Bersal del Carmen Villegas Camposeco for their valuable technical and analytical support. We are grateful to Jorge A. Gómez Valdés, Stephan Schiffels, Choongwon Jeong, Arturo Pacho Varela, Kathrin Nägele, Vanessa Villalba, Cosimo Posth, Maria Spyrou, Felix Key, and Maïté Rivollat for insightful comments and helpful discussions, which are part of the finished version of this manuscript. We acknowledge the Council of Archaeology, National Institute of Anthropology and History (INAH) (Mexico City, Mexico) for the permit granted to analyze the individuals from the San José de los Naturales collection (project: *Identificación de los sitios de origen de los individuos con limado dental característico de grupos africanos el Hospital Real de San José de los Naturales a través de análisis genético molecular*; official notice number: 401.B(4)19.2015/36/1026; final report approval: 401.1S.3-2020/0112).

was supported by the University Research Priority Program of Evolution in Action of the University of Zurich. This work was financially supported by the Max Planck Society.

AUTHOR CONTRIBUTIONS

R.B., D.I.H.-Z., N.B.-F., V.A.-A., A.H., D.K., L.M.-M., and J.K. conceived the project. R.B., D.I.H.-Z., N.B.-F., and L.M.-M. performed the registration and photographic recording. R.B., D.I.H.-Z., E.A.N., A.C.Z.-H., and N.B.-F. performed the osteological and paleopathological analyses. R.B., L.M.-M., and J.K. performed the sampling. R.B. performed the lab work. A.I. contributed to the design of the HLA capture. A.H. developed the hepatitis B virus in-solution capture design. P. Roberts and P. Ramallo performed the isotopes analyses. R.B., T.C.L., C.B., and J.K. analyzed the population genetics data. A.K., A.K.L., A.H., D.K., and J.K. assembled, analyzed, and discussed the hepatitis B virus genome and the *Treponema pallidum* sub. *pertenue* genome. All authors discussed the results. R.B., T.C.L., D.I.H.-Z., E.A.N., L.M.-M., and J.K. wrote the paper with contributions from all authors. All authors read and approved the final version. R.B., D.I.H.-Z., N.B.-F., and L.M.-M. wrote and submitted the final report (in Spanish) for the Council of Archaeology (INAH) based on the present manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 17, 2020
Revised: March 3, 2020
Accepted: April 1, 2020
Published: April 30, 2020

REFERENCES

1. Lockhart, J., and Schwartz, S. (1983). Early Latin America (Cambridge University Press).
2. Acuña-Alonzo, V. (2005). La contribución genética africana a las poblaciones contemporáneas. PhD thesis (Escuela Nacional de Antropología e Historia).
3. Manning, P. (1993). Migrations of Africans to the Americas: the impact on Africans, Africa, and the New World. Hist. Teacher *26*, 279–296.
4. Lovejoy, P.E. (2011). Esclavitud y comercio esclavista en el África Occidental: investigaciones en curso. In Debates Históricos Contemporáneos: Africanos y Afrodescendientes en México y Centroamérica, M.E. Velázquez, ed. (Centro de Estudios Mexicanos y Centroamericanos), pp. 35–57.
5. Barquera, R., and Acuña-Alonzo, V. (2012). The African colonial migration into Mexico: history and biological consequences. In Causes and Consequences of Human Migration: An Evolutionary Perspective, M.H. Crawford, and B.C. Campbell, eds. (Cambridge University Press), pp. 201–223.
6. Bernal Felipe, N., and Barquera Lozano, R. (2017). Las raíces africanas en México: perspectivas desde la antropología física. In Antropología Física: Disciplina Bio-psico-social, L. González Quintero, and A. Barragán Solís, eds. (Instituto Nacional de Antropología e Historia), pp. 227–250.
7. Aguirre Beltrán, G. (1972). La Población Negra en México: Estudio Etnohistórico, Second Edition (Fondo de Cultura Económica).
8. Lisker, R., Loria, A., and Cordova, M.S. (1965). Studies on several genetic hematological traits of the mexican population. 8. Hemoglobin s, glucose-6-phosphate dehydrogenase deficiency, and other characteristics in a malarial region. Am. J. Hum. Genet. *17*, 179–187.
9. Vågene, Å.J., Herbig, A., Campana, M.G., Robles García, N.M., Warinner, C., Sabin, S., Spyrou, M.A., Andrades Valtueña, A., Huson, D., Tuross, N., et al. (2018). Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. Nat. Ecol. Evol. *2*, 520–528.
10. Marr, J.S., and Kiracofe, J.B. (2000). Was the huey cocoliztli a haemorrhagic fever? Med. Hist. *44*, 341–362.
11. Burns, J.N., Acuña-Soto, R., and Stahle, D.W. (2014). Drought and epidemic typhus, central Mexico, 1655-1918. Emerg. Infect. Dis. *20*, 442–447.
12. Wang, S., Ray, N., Rojas, W., Parra, M.V., Bedoya, G., Gallo, C., Poletti, G., Mazzotti, G., Hill, K., Hurtado, A.M., et al. (2008). Geographic patterns of genome admixture in Latin American Mestizos. PLoS Genet. *4*, e1000037.
13. Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., Fuentes, M., Pizarro, M., Everardo, P., de Avila, F., et al. (2014). Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. PLoS Genet. *10*, e1004572.
14. Harvey, S.P. (2016). Ideas of race in early America. Oxford Res. Encycl. https://oxfordre.com/americanhistory/view/10.1093/acrefore/9780199329175.001.0001/acrefore-9780199329175-e-262.
15. Zavala, S. (1968). Los Esclavos Indios en Nueva España (El Colegio de México).
16. Schroeder, H., Ávila-Arcos, M.C., Malaspinas, A.-S., Poznik, G.D., Sandoval-Velasco, M., Carpenter, M.L., Moreno-Mayar, J.V., Sikora, M., Johnson, P.L.F., Allentoft, M.E., et al. (2015). Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. Proc. Natl. Acad. Sci. USA *112*, 3669–3673.
17. Bourne, J.K., Jr. (2018). Digging for the life stories of long-forgotten slaves. Natl. Geogr. Mag. https://www.nationalgeographic.com/culture/2018/12/charleston-gullah-dna-anson-slave-burials/.
18. Schablitsky, J.M., Witt, K.E., Ramos Madrigal, J., Ellegaard, M.R., Malhi, R.S., and Schroeder, H. (2019). Ancient DNA analysis of a nineteenth century tobacco pipe from a Maryland slave quarter. J. Archaeol. Sci. *105*, 11–18.
19. Fortes-Lima, C., Gessain, A., Ruiz-Linares, A., Bortolini, M.C., Migot-Nabias, F., Bellis, G., Moreno-Mayar, J.V., Restrepo, B.N., Rojas, W., Avendaño-Tamayo, E., et al. (2017). Genome-wide ancestry and demographic history of African-descendant maroon communities from French Guiana and Suriname. Am. J. Hum. Genet. *101*, 725–736.
20. Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J.L., Byrnes, J.K., Gignoux, C.R., Ortiz-Tello, P.A., Martínez, R.J., Hedges, D.J., Morris, R.W., et al. (2013). Reconstructing the population genetic history of the Caribbean. PLoS Genet. *9*, e1003925.
21. Cabrera Torres, J.J., García Martínez, M., and de los, Á. (1998). Utilización, modificación y reúso de los espacios del edificio sede del Hospital Real de San José de los Naturales. PhD thesis (Escuela Nacional de Antropología e Historia).
22. Zedillo, A. (1984). Historia de un Hospital: El Hospital Real de Naturales, First Edition (Editorial del Instituto Mexicano del Seguro Social).
23. López Wario, L.A., Meza Peñaloza, A., and Báez-Molgado, S. (1996). Una muerte violenta en el virreinato (el caso del esqueleto 150 de la línea 8 del Metro, México, D.F.). Rev. la Coord. Nac. Arqueol. del Inst. Nac. Antropol. e Hist. *Ene-jun*, 111–114.
24. Del Castillo Chávez, O. (2000). Condiciones de vida y salud de una muestra poblacional de la Ciudad de México en la época colonial. PhD thesis (Escuela Nacional de Antropología e Historia).
25. Lagunas Rodríguez, Z., and Karam Tapia, C.E. (2003). Cráneos africanos de la época colonial con mutilación dentaria, procedentes del ex Hospital Real de San José de los Naturales de la Ciudad de México, D.F. Estud. Antropol. Biologica (Santiago) *11*, 967–981.
26. Hernández López, P.E., and Negrete Gutiérrez, S.S. (2012). ¿Realmente eran indios? Afinidad biológica entre las personas atendidas en el Hospital Real San José de los Naturales, siglos XVI-XVIII. PhD thesis (Escuela Nacional de Antropología e Historia).
27. Meza, A. (2013). Presencia africana en el cementerio del Hospital Real de San José de los Naturales. Arqueol. Mex. *21*, 40–44.

28. Báez Molgado, S., and Meza-Peñaloza, A. (1995). Análisis de los Restos Óseos del Hospital Real de San José de los Naturales (Individuos Completos). Report. Archivo Técnico de la Dirección de Salvamento Arqueológico (DSA), Instituto Nacional de Antropología e Historia (INAH).

29. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al.; International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

30. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.

31. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.

32. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC Bioinformatics *15*, 356.

33. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. *16*, 224.

34. Wright, L.E., and Schwarcz, H.P. (1998). Stable carbon and oxygen isotopes in human tooth enamel: identifying breastfeeding and weaning in prehistory. Am. J. Phys. Anthropol. *106*, 1–18.

35. Sealy, J.C., van der Merwe, N.J., Thorp, J.A.L., and Lanham, J.L. (1987). Nitrogen isotopic ecology in southern Africa: implications for environmental and dietary tracing. Geochim. Cosmochim. Acta *51*, 2707–2717.

36. Loftus, E., Roberts, P., and Lee-Thorp, J.A. (2016). An isotopic generation: four decades of stable isotope analysis in African archaeology. Azania *51*, 88–114.

37. Walker, P.L., Bathurst, R.R., Richman, R., Gjerdrum, T., and Andrushko, V.A. (2009). The causes of porotic hyperostosis and cribra orbitalia: a reappraisal of the iron-deficiency-anemia hypothesis. Am. J. Phys. Anthropol. *139*, 109–125.

38. Goodman, A.H. (1994). Cartesian reductionism and vulgar adaptationism: issues in the interpretation of nutritional status in prehistory. In Paleonutrition: The Diet of Prehistoric Americans, K. Sobolik, ed. (Southern Illinois University), pp. 163–177.

39. Larsen, C.S. (1997). Bioarchaeology: Interpreting Behavior from the Human Skeleton (Cambridge University).

40. Klaus, H.D., and Tam, M.E. (2009). Contact in the Andes: bioarchaeology of systemic stress in colonial Mórrope, Peru. Am. J. Phys. Anthropol. *138*, 356–368.

41. Wasterlain, S.N., Neves, M.J., and Ferreira, M.T. (2016). Dental modifications in a skeletal sample of enslaved Africans found at Lagos (Portugal). Int. J. Osteoarchaeol. *26*, 621–632.

42. Alves, R.V., Garcia, S.J., Marques, A., and Wasterlain, S.N. (2016). Osteological analysis of a skeleton with intentional dental modifications, exhumed from Largo do Carmo (17th – 18th centuries), Lisbon. Antropol. Port. *32*, 61–75.

43. Lignitz, H. (1919). Die künstlichen Zahnverstümmlungen in Afrika im Lichte der Kulturkreisforschung. Anthropos *14/15*, 891–943.

44. Rosa, A., and Brehem, A. (2011). African human mtDNA phylogeography at-a-glance. J. Anthropol. Sci. *89*, 25–58.

45. Brehm, A., Pereira, L., Bandelt, H.J., Prata, M.J., and Amorim, A. (2002). Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. Ann. Hum. Genet. *66*, 49–60.

46. González, A.M., Cabrera, V.M., Larruga, J.M., Tounkara, A., Noumsi, G., Thomas, B.N., and Moulds, J.M. (2006). Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other Western Africa populations. Ann. Hum. Genet. *70*, 631–657.

47. Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M.J., and Amorim, A. (2001). Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. Ann. Hum. Genet. *65*, 439–458.

48. Salas, A., Richards, M., De la Fe, T., Lareu, M.-V., Sobrino, B., Sánchez-Diz, P., Macaulay, V., and Carracedo, A. (2002). The making of the African mtDNA landscape. Am. J. Hum. Genet. *71*, 1082–1111.

49. Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonné-Tamir, B., Bertranpetit, J., Francalacci, P., et al. (2000). Y chromosome sequence variation and the history of human populations. Nat. Genet. *26*, 358–361.

50. Vallone, P.M., and Butler, J.M. (2004). Y-SNP typing of U.S. African American and Caucasian samples using allele-specific hybridization and primer extension. J. Forensic Sci. *49*, 723–732.

51. Schuster, S.C., Miller, W., Ratan, A., Tomsho, L.P., Giardine, B., Kasson, L.R., Harris, R.S., Petersen, D.C., Zhao, F., Qi, J., et al. (2010). Complete Khoisan and Bantu genomes from southern Africa. Nature *463*, 943–947.

52. Trombetta, B., Cruciani, F., Sellitto, D., and Scozzari, R. (2011). A new topology of the human Y chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms. PLoS ONE *6*, e16073.

53. Martínez-Cortés, G., Salazar-Flores, J., Haro-Guerrero, J., Rubi-Castellanos, R., Velarde-Félix, J.S., Muñoz-Valle, J.F., López-Casamichana, M., Carrillo-Tapia, E., Canseco-Avila, L.M., Bravi, C.M., et al. (2013). Maternal admixture and population structure in Mexican-Mestizos based on mtDNA haplogroups. Am. J. Phys. Anthropol. *151*, 526–537.

54. Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher, O. (2014). OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics *30*, 3310–3316.

55. Assane, A.A.A., Fabricio-Silva, G.M., Cardoso-Oliveira, J., Mabunda, N.E.J., Sousa, A.M., Jani, I.V., Ferreira, O.C., Jr., and Porto, L.C.M.S. (2010). Human leukocyte antigen-A, -B, and -DRB1 allele and haplotype frequencies in the Mozambican population: a blood donor-based population study. Hum. Immunol. *71*, 1027–1032.

56. Paximadis, M., Mathebula, T.Y., Gentle, N.L., Vardas, E., Colvin, M., Gray, C.M., Tiemessen, C.T., and Puren, A. (2012). Human leukocyte antigen class I (A, B, C) and II (DRB1) diversity in the black and Caucasian South African population. Hum. Immunol. *73*, 80–92.

57. Arlehamn, C.S.L., Copin, R., Leary, S., Mack, S.J., Phillips, E., Mallal, S., Sette, A., Blatner, G., Siefers, H., and Ernst, J.D.; TBRU-ASTRa Study Team (2017). Sequence-based HLA-A, B, C, DP, DQ, and DR typing of 100 Luo infants from the Boro area of Nyanza Province, Kenya. Hum. Immunol. *78*, 325–326.

58. Grifoni, A., Sidney, J., Carpenter, C., Phillips, E., Mallal, S., Scriba, T.J., Sette, A., and Lindestam Arlehamn, C.S. (2018). Sequence-based HLA-A, B, C, DP, DQ, and DR typing of 159 individuals from the Worcester region of the Western Cape province of South Africa. Hum. Immunol. *79*, 143–144.

59. Maiers, M., Gragert, L., and Klitz, W. (2007). High-resolution HLA alleles and haplotypes in the United States population. Hum. Immunol. *68*, 779–788.

60. Nunes, K., Piovezan, B., Torres, M.A., Pontes, G.N., Kimura, L., Carnavalli, J.E.P., Mingroni Netto, R.C., Moraes, M.E., and Meyer, D. (2016). Population variation of HLA genes in rural communities in Brazil, the Quilombos from the Vale do Ribeira, São Paulo - Brazil. Hum. Immunol. *77*, 447–448.

61. Tang, J., Naik, E., Costello, C., Karita, E., Rivers, C., Allen, S., and Kaslow, R.A. (2000). Characteristics of HLA class I and class II polymorphisms in Rwandan women. Exp. Clin. Immunogenet. *17*, 185–198.

62. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445–449.

63. Price, T.D., Tiesler, V., and Burton, J.H. (2006). Early African Diaspora in colonial Campeche, Mexico: strontium isotopic evidence. Am. J. Phys. Anthropol. *130*, 485–490.

64. Price, T.D., Burton, J.H., Cucina, A., Zabala, P., Frei, R., Tykot, R.H., and Tiesler, V. (2012). Isotopic studies of human skeletal remains from a sixteenth to seventeenth century AD churchyard in Campeche, Mexico. Curr. Anthropol. *53*, 396–433.

65. Hodell, D.A., Quinn, R.L., Brenner, M., and Kamenov, G. (2004). Spatial variation of strontium isotopes ($^{87}Sr/^{86}Sr$) in the Maya region: a tool for tracking ancient human migration. J. Archaeol. Sci. *31*, 585–601.

66. Juarez, C.A. (2008). Strontium and geolocation, the pathway to identification for deceased undocumented mexican border-crossers: a preliminary report. J. Forensic Sci. *53*, 46–49.

67. Schroeder, H., O'Connell, T.C., Evans, J.A., Shuler, K.A., and Hedges, R.E.M. (2009). Trans-Atlantic slavery: isotopic evidence for forced migration to Barbados. Am. J. Phys. Anthropol. *139*, 547–557.

68. Hübler, R., Key, F.M., Warinner, C., Bos, K.I., Krause, J., and Herbig, A. (2019). HOPS: automated detection and authentication of pathogen DNA in archaeological remains. Genome Biol. *20*, 280.

69. Kramvis, A. (2014). Genotypes and genetic variability of hepatitis B virus. Intervirology *57*, 141–150.

70. Pourkarim, M.R., Amini-Bavil-Olyaee, S., Kurbanov, F., Van Ranst, M., and Tacke, F. (2014). Molecular identification of hepatitis B virus genotypes/subgenotypes: revised classification hurdles and updated resolutions. World J. Gastroenterol. *20*, 7152–7168.

71. Schuenemann, V.J., Kumar Lankapalli, A., Barquera, R., Nelson, E.A., Iraíz Hernández, D., Acuña Alonzo, V., Bos, K.I.K.I., Márquez Morfín, L., Herbig, A., and Krause, J. (2018). Historic Treponema pallidum genomes from Colonial Mexico retrieved from archaeological remains. PLoS Negl. Trop. Dis. *12*, e0006447.

72. Hurston, Z.N. (2018). Barracoon. The Story of the Last "Black Cargo,", First Edition (Amistad Press/HarperCollins Publishers).

73. Corona Paredes, O. (2004). Informe Final del Rescate Arqueológico en el ex Templo de Corpus Christi. Report. Dirección de Salvamento Arqueológico (INAH).

74. Green, L.D., Derr, J.N., and Knight, A. (2000). mtDNA affinities of the peoples of North-Central Mexico. Am. J. Hum. Genet. *66*, 989–998.

75. Guardado-Estrada, M., Juarez-Torres, E., Medina-Martinez, I., Wegier, A., Macías, A., Gomez, G., Cruz-Talonia, F., Roman-Bassaure, E., Piñero, D., Kofman-Alfaro, S., and Berumen, J. (2009). A great diversity of Amerindian mitochondrial DNA ancestry is present in the Mexican mestizo population. J. Hum. Genet. *54*, 695–705.

76. Barquera, R., Hernandez-Zaragoza, D.I., and Bravo Acevedo, A. (2018). The immunogenetic diversity of the HLA system in Mexico correlates with underlying population genetic structure. The Allele Frequency Net Database. http://allelefrequencies.net/hla6006a.asp?page=1&hla_locus=&hla_locus_type=Classical&hla_allele1=&hla_allele2=&hla_selection=&hla_pop_selection=&hla_population=&hla_country=Mexico&hla_dataset=&hla_region=&hla_ethnic=&hla_study=Other&hla_sample_size=&hla_s.

77. Li, Z.-K., Nie, J.-J., Li, J., and Zhuang, H. (2013). The effect of HLA on immunological response to hepatitis B vaccine in healthy people: a meta-analysis. Vaccine *31*, 4355–4361.

78. Yoon, J.H., Shin, S., In, Jw., Chang, J.Y., Song, E.Y., and Roh, E.Y. (2014). Association of HLA alleles with the responsiveness to hepatitis B virus vaccination in Korean infants. Vaccine *32*, 5638–5644.

79. Nishida, N., Sawai, H., Kashiwase, K., Minami, M., Sugiyama, M., Seto, W.-K., Yuen, M.-F., Posuwan, N., Poovorawan, Y., Ahn, S.H., et al. (2014). New susceptibility and resistance HLA-DP alleles to HBV-related diseases identified by a trans-ethnic association study in Asia. PLoS ONE *9*, e86449.

80. Jiang, H.-W., Tian, H.-Q., Liu, H., Li, N., Zhao, Y., and Zhang, F.-R. (2011). Association of the HLA-DRB1 locus with syphilis in a Chinese population. Int. J. Infect. Dis. *15*, e342–e345.

81. Aguirre Beltrán, G. (1944). The slave trade in Mexico. Hisp. Am. Hist. Rev. *24*, 412–431.

82. Lavanchy, D., and Kane, M. (2016). Global epidemiology of hepatitis B virus infection BT - hepatitis B virus in human diseases. In Hepatitis B Virus in Human Diseases, Y.-F. Liaw, and F. Zoulim, eds. (Springer International Publishing), pp. 187–203.

83. Krause-Kyora, B., Susat, J., Key, F.M., Kühnert, D., Bosse, E., Immel, A., Rinne, C., Kornell, S.-C., Yepes, D., Franzenburg, S., et al. (2018). Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. eLife *7*, e36666.

84. Mühlemann, B., Jones, T.C., Damgaard, P.B., Allentoft, M.E., Shevnina, I., Logvin, A., Usmanova, E., Panyushkina, I.P., Boldgiv, B., Bazartseren, T., et al. (2018). Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature *557*, 418–423.

85. Andernach, I.E., Nolte, C., Pape, J.W., and Muller, C.P. (2009). Slave trade and hepatitis B virus genotypes and subgenotypes in Haiti and Africa. Emerg. Infect. Dis. *15*, 1222–1228.

86. Morfín, M., and Espinoza, L.H.P. (2006). Salud y Sociedad en el México Prehispánico y Colonial (Instituto Nacional de Antropología e Historia/Escuela Nacional de Antropología e Historia).

87. Márquez Morfín, L., and Meza Manzanilla, M. (2015). Sífilis en la Ciudad de México: análisis osteopatológico. Cuicuilco *22*, 89–126.

88. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.

89. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

90. Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biol. *17*, 60.

91. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

92. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684.

93. Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc. Natl. Acad. Sci. USA *111*, 2229–2234.

94. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

95. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

96. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.

97. Venegas Ramírez, C. (1973). Régimen Hospitalario para Indios en la Nueva España (Secretaría de Educación Pública/Instituto Nacional de Antropología e Historia).

98. Suárez, M. (1988). Hospitales y sociedad en la Ciudad de México en el siglo XVI. PhD thesis (Universidad Autónoma Metropolitana-Azcapotzalco).

99. Muriel, J. (1990). Hospital Real de Sanct Joseph de los Naturales, México, D. F. In Hospitales de la Nueva España. T. 1. Fundaciones del siglo XVI (Universidad Nacional Autónoma de México—Instituto de Investigaciones Históricas/Cruz Roja Mexicana), pp. 127–148.

100. McCaa, R. (1995). Spanish and Nahuatl views on smallpox and demographic catastrophe in Mexico. J. Interdiscip. Hist. *25*, 397–431.

101. Ruíz Albarrán, P., and del, C. (2012). Estudios de variabilidad biológica en la colección esquelética Hospital Real de Naturales. Un acercamiento a

través de la técnica de Morfometría Geométrica. PhD thesis (Escuela Nacional de Antropología e Historia).

102. Fierros Millán, J. (2009). El Hospital Real de Naturales (1701-1741), un hospital disputado de la capital novohispana. PhD thesis (Escuela Nacional de Antropología e Historia).
103. Garve, M. (2011). Rituelle Deformierungen der Zähne und deren Einfluss auf das orofaziale System bei Naturvölkern am Beispiel der Bench in Südwest- Äthiopien. PhD thesis (Universitätsmedizin der Ernst-Moritz-Arndt-Universität Greifswald).
104. Goodman, A.H., and Martin, D.L. (2002). Reconstructing health profiles from skeletal remains. In The Backbone of History. Health and Nutricion of the Western Hemisphere, R.H. Steckel, and J.C. Rose, eds. (Cambridge University), pp. 11–60.
105. Ortner, D.J. (2003). Identification of Pathological Conditions in Human Skeletal Remains, Second Edition (Academic).
106. Buikstra, J.E., and Ubelaker, D.H. (1994). Standards for Data Collection from Human Skeletal Remains: Proceedings of a Seminar at the Field Museum of Natural History (Arkansas Archeological Survey).
107. Hernández Espinoza, P.O., and Rodríguez Lagunas, Z. (2015). Manual de Osteología, Third Edition (Instituto Nacional de Antropología e Historia/ Escuela Nacional de Antropología e Historia).
108. Reichart, P.A., Creutz, U., and Scheifele, C. (2008). Dental mutilations and associated alveolar bone pathology in African skulls of the anthropological skull collection, Charité, Berlin. J. Oral Pathol. Med. *37*, 50–55.
109. Hefner, J.T. (2009). Cranial nonmetric variation and estimating ancestry. J. Forensic Sci. *54*, 985–995.
110. Coelho, J. d'Oliveira, and Navega, D. (2011). Cranial nonmetric traits ancestry estimation. Osteomics. http://osteomics.com/hefneR/.
111. Villegas Camposeco, B., and del, C. (2018). ¿Tu origen, es mi origen? Frecuencia y variación de rasgos no-métricos craneales en población mexicana contemporánea. PhD thesis (Escuela Nacional de Antropología e Historia).
112. Bruzek, J. (2002). A method for visual determination of sex, using the human hip bone. Am. J. Phys. Anthropol. *117*, 157–168.
113. DiGangi, E.A., Bethard, J.D., Kimmerle, E.H., and Konigsberg, L.W. (2009). A new method for estimating age-at-death from the first rib. Am. J. Phys. Anthropol. *138*, 164–176.
114. McKern, T.W., and Stewart, T.D. (1957). Skeletal Age Changes in Young American Males: Analysed from the Standpoint of Age Identification (Headquarters, Quartermaster Research & Development Command, Quartermaster Research & Development Center, Environmental Protection Research Division).
115. Lovejoy, C.O., Meindl, R.S., Pryzbeck, T.R., and Mensforth, R.P. (1985). Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death. Am. J. Phys. Anthropol. *68*, 15–28.
116. Henríquez, M., and Arriaza, B. (2013). Distribución y frecuencia de nódulos de Schmörl en la columna vertebral de poblaciones prehispánicas de Arica: ¿Indicadores de la carga laboral? Rev. Antropol. Chil. *45*, 311–319.
117. Phenice, T.W. (1969). A newly developed visual method of sexing the os pubis. Am. J. Phys. Anthropol. *30*, 297–301.
118. Ubelaker, D.H. (1991). Perimortem and postmortem modifications of human bone: lessons from forensic anthropology. Anthropologie *29*, 171–174.
119. Ascough, P.L., Cook, G.T., and Dugmore, A. (2005). Methodological approaches to determining the marine radiocarbon reservoir effect. Prog. Phys. Geogr. *29*, 532–547.
120. Bronk Ramsey, C. (2008). Radiocarbon dating: revolutions in understanding. Archaeometry *50*, 249–275.
121. AlQahtani, S.J., Hector, M.P., and Liversidge, H.M. (2010). Brief communication: the London atlas of human tooth development and eruption. Am. J. Phys. Anthropol. *142*, 481–490.
122. Logan, W.H.G., and Kronfeld, R. (1933). Development of the human jaws and surrounding structures from birth to the age of fifteen years. J. Am. Dent. Assoc. *20*, 379–428.
123. Richards, M.P., and Hedges, R.E.M. (1999). Stable isotope evidence for similarities in the types of marine foods used by late Mesolithic humans at sites along the Atlantic coast of Europe. J. Archaeol. Sci. *26*, 717–722.
124. van Klinken, G.J. (1999). Bone collagen quality indicators for palaeodietary and radiocarbon measurements. J. Archaeol. Sci. *26*, 687–695.
125. DeNiro, M.J. (1985). Postmortem preservation and alteration of *in vivo* bone collagen isotope ratios in relation to palaeodietary reconstruction. Nature *317*, 806–809.
126. Roberts, P., Fernandes, R., Craig, O.E., Larsen, T., Lucquin, A., Swift, J., and Zech, J. (2018). Calling all archaeologists: guidelines for terminology, methodology, data handling, and reporting when undertaking and reviewing stable isotope applications in archaeology. Rapid Commun. Mass Spectrom. *32*, 361–372.
127. Ambrose, S.H., and Norr, L. (1993). Experimental evidence for the relationship of the carbon isotope ratios of whole diet and dietary protein to those of bone collagen and carbonate. In Prehistoric Human Bone: Archaeology at the Molecular Level, J.B. Lambert, and G. Grupe, eds. (Springer Berlin Heidelberg), pp. 1–37.
128. Smith, B.N., and Epstein, S. (1971). Two categories of c/c ratios for higher plants. Plant Physiol. *47*, 380–384.
129. Farquhar, G.D., Ehleringer, J.R., and Hubick, K.T. (1989). Carbon isotope discrimination and photosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. *40*, 503–537.
130. Tieszen, L.L. (1991). Natural variations in the carbon isotope values of plants: implications for archaeology, ecology, and paleoecology. J. Archaeol. Sci. *18*, 227–248.
131. Deniro, M.J., and Epstein, S. (1981). Influence of diet on the distribution of nitrogen isotopes in animals. Geochim. Cosmochim. Acta *45*, 341–351.
132. Ambrose, S.H. (1991). Effects of diet, climate and physiology on nitrogen isotope abundances in terrestrial foodwebs. J. Archaeol. Sci. *18*, 293–317.
133. Hedges, R.E.M., and Reynard, L.M. (2007). Nitrogen isotopes and the trophic level of humans in archaeology. J. Archaeol. Sci. *34*, 1240–1251.
134. Schoeninger, M.J., and DeNiro, M.J. (1984). Nitrogen and carbon isotopic composition of bone collagen from marine and terrestrial animals. Geochim. Cosmochim. Acta *48*, 625–639.
135. Dufour, E., Bocherens, H., and Mariotti, A. (1999). Palaeodietary implications of isotopic variability in Eurasian lacustrine fish. J. Archaeol. Sci. *26*, 617–627.
136. Handler, J.S. (2002). Survivors of the middle passage: life histories of enslaved Africans in British America. Slavery Abol. *23*, 25–56.
137. Cox, G., Sealy, J., Schrire, C., and Morris, A. (2001). Stable carbon and nitrogen isotopic analyses of the underclass at the colonial Cape of Good Hope in the eighteenth and nineteenth centuries. World Archaeol. *33*, 73–97.
138. Carney, J.A. (2001). African rice in the Columbian exchange. J. Afr. Hist. *42*, 377–396.
139. Stahl, A.B. (1999). The archaeology of global encounters viewed from Banda, Ghana. Afr. Archaeol. Rev. *16*, 5–81.
140. Bastos, M.Q.R., Santos, R.V., de Souza, S.M.F.M., Rodrigues-Carvalho, C., Tykot, R.H., Cook, D.C., and Santos, R.V. (2016). Isotopic study of geographic origins and diet of enslaved Africans buried in two Brazilian cemeteries. J. Archaeol. Sci. *70*, 82–90.
141. Laffoon, J.E., Espersen, R., and Mickleburgh, H.L. (2018). The life history of an enslaved African: multiple isotope evidence for forced childhood migration from Africa to the Caribbean and associated dietary change. Archaeometry *60*, 350–365.
142. Fernandes, R., Rinne, C., Nadeau, M.-J., and Grootes, P. (2016). Towards the use of radiocarbon as a dietary proxy: establishing a first wide-ranging radiocarbon reservoir effects baseline for Germany. Environ. Archaeol. *21*, 285–294.

143. Cook, G.T., and van der Plicht, J. (2006). Radiocarbon dating: conventional method. In Encyclopedia of Quaternary Science, S. Elias, ed. (Elsevier), pp. 2899–2911.

144. Keaveney, E.M., and Reimer, P.J. (2012). Understanding the variability in freshwater radiocarbon reservoir offsets: a cautionary tale. J. Archaeol. Sci. *39*, 1306–1316.

145. Jull, A.J.T., Burr, G.S., and Hodgins, G.W.L. (2013). Radiocarbon dating, reservoir effects, and calibration. Quat. Int. *299*, 64–71.

146. Rohland, N., and Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. Nat. Protoc. *2*, 1756–1762.

147. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. USA *110*, 15758–15763.

148. Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., and Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. USA *104*, 14616–14621.

149. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, pdb.prot5448.

150. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. Philos. Trans. R. Soc. B Biol. Sci. *370*, 20130624.

151. Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., et al. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat. Biotechnol. *27*, 182–189.

152. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.

153. Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., and Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. Proc. Natl. Acad. Sci. USA *110*, 2223–2227.

154. Fu, Q., Hajdinjak, M., Moldovan, O.T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. Nature *524*, 216–219.

155. Ghodsi, M., Liu, B., and Pop, M. (2011). DNACLUST: accurate and efficient clustering of phylogenetic marker genes. BMC Bioinformatics *12*, 271.

156. Immel, A., Key, F.M., Szolek, A., Barquera, R., Robinson, M.K., Spyrou, M.A., Susat, J., Krause-Kyora, B., Bos, K.I., Forrest, S., et al. (2019). Genomic DNA from Late Medieval plague victims suggests effect of Yersinia pestis on human immunity genes. bioRxiv.

157. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44* (D1), D733–D745.

158. Robinson, J., Halliwell, J.A., Hayhurst, J.D., Flicek, P., Parham, P., and Marsh, S.G.E. (2015). The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. *43*, D423–D431.

159. Holdsworth, R., Hurley, C.K., Marsh, S.G.E., Lau, M., Noreen, H.J., Kempenich, J.H., Setterholm, M., and Maiers, M. (2009). The HLA dictionary 2008: a summary of HLA-A, -B, -C, -DRB1/3/4/5, and -DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR, and -DQ antigens. Tissue Antigens *73*, 95–170.

160. Weese, D., Holtgrewe, M., and Reinert, K. (2012). RazerS 3: faster, fully sensitive read mapping. Bioinformatics *28*, 2592–2599.

161. Schiffels, S. (2018). sequenceTools. https://github.com/stschiff/sequenceTools.git.

162. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. *23*, 147.

163. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J.J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. *44* (W1), W58–W63.

164. Vianello, D., Sevini, F., Castellani, G., Lomartire, L., Capri, M., and Franceschi, C. (2013). HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. Hum. Mutat. *34*, 1189–1194.

165. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

166. Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics *26*, 419–420.

167. Banks, R. (2018). International Society of Genetic Genealogy database (ISOGG) Y-DNA. Y-DNA Haplogr. Tree 2018 Version 13.174. https://isogg.org/tree/.

168. Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., Hammer, M., Bustamante, C.D., and Ostrer, H. (2010). Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. Proc. Natl. Acad. Sci. USA *107* (*Suppl 2*), 8954–8961.

169. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

170. Alexander, D.H., Novembre, J., and Lange, K. (2016). ADMIXTURE. https://www.genetics.ucla.edu/software/admixture/download.html.

171. Moreno-Estrada, A., Gignoux, C.R., Fernández-López, J.C., Zakharia, F., Sikora, M., Contreras, A.V., Acuña-Alonzo, V., Sandoval, K., Eng, C., Romero-Hidalgo, S., et al. (2014). Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. Science *344*, 1280–1285.

172. Patterson, N. (2012). DReichLab/AdmixTools. https://github.com/DReichLab/AdmixTools.

173. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. *17*, 540–552.

174. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

175. Herbig, A., Maixner, F., Bos, K.I., Zink, A., Krause, J., and Huson, D.H. (2016). MALT: fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. bioRxiv. https://doi.org/10.1101/050559.

176. Broad Institute (2019). Picard Toolkit (GitHub Repos).

177. Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., et al. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature *514*, 494–497.

178. Didelot, X., and Wilson, D.J. (2015). ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput. Biol. *11*, e1004041.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Ancient individual	This study/ San José de los Naturales archaeological site	SJN001/ *ML8 SL 150*
Ancient individual	This study/ San José de los Naturales archaeological site	SJN002/ *ML8 San José 214*
Ancient individual	This study/ San José de los Naturales archaeological site	SJN003/ *ML8 SLU9B 296*
Chemicals, Peptides, and Recombinant Proteins		
2x HI-RPM hybridization buffer	Agilent Technologies	Cat# 5190-0403
Herculase II Fusion DNA Polymerase	Agilent Technologies	Cat# 600679
Pfu Turbo Cx Hotstart DNA Polymerase	Agilent Technologies	Cat# 600412
D1000 ScreenTapes	Agilent Technologies	Cat# 5067-5582
D1000 Reagents	Agilent Technologies	Cat# 5067-5583
0.5 M EDTA pH 8.0	BioExpress	Cat# E177
Sera-Mag Magnetic Speed-beads Carboxylate-Modified (1 mm, 3EDAC/PA5)	GE LifeScience	Cat# 65152105050250
Bst DNA Polymerase2.0, large frag.	New England Biolabs	Cat# M0537
UGI	New England Biolabs	Cat# M0281
USER enzyme	New England Biolabs	Cat# M5505
PE buffer concentrate	QiaGEN	Cat# 19065
1 M Tris-HCl pH 8.0	Sigma Aldrich	Cat# AM9856
1M NaOH	Sigma Aldrich	Cat# 71463
20% SDS	Sigma Aldrich	Cat# 5030
3M Sodium Acetate (pH 5.2)	Sigma Aldrich	Cat# S7899
5M NaCl	Sigma Aldrich	Cat# S5150
Ethanol	Sigma Aldrich	Cat# E7023
Guanidine hydrochloride	Sigma Aldrich	Cat# G3272
Isopropanol	Sigma Aldrich	Cat# 650447
PEG-8000	Sigma Aldrich	Cat# 89510
Proteinase K	Sigma Aldrich	Cat# P6556
Tween-20	Sigma Aldrich	Cat# P9416
Water	Sigma Aldrich	Cat# W4502
10x Buffer Tango	Thermo Fisher Scientific	Cat# BY5
50x Denhardt's solution	Thermo Fisher Scientific	Cat# 750018
ATP	Thermo Fisher Scientific	Cat# R0441
dNTP Mix	Thermo Fisher Scientific	Cat# R1121
Dynabeads MyOne Streptavidin T1	Thermo Fisher Scientific	Cat# 65602
DyNAmo HS SYBR Green qPCR Kit	Thermo Fisher Scientific	Cat# F410L
GeneAmp 10x PCR Gold Buffer	Thermo Fisher Scientific	Cat# 4379874
Human Cot-I DNA	Thermo Fisher Scientific	Cat# 15279011
Salmon sperm DNA	Thermo Fisher Scientific	Cat# 15632-011
SSC Buffer (20x)	Thermo Fisher Scientific	Cat# AM9770
T4 DNA Ligase	Thermo Fisher Scientific	Cat# EL0011
T4 DNA Polymerase	Thermo Fisher Scientific	Cat# EP0062
T4 Polynucleotide Kinase	Thermo Fisher Scientific	Cat# EK0032

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Acetone, certified ACS	VWR	Cat# BDH1101-4LP
Dichloromethane, certified ACS	VWR	Cat# EMD-DX0835-3
Hydrochloric acid, 6N, 0.5N & 0.01N	VWR	Cat# EMD-HX0603-3
Critical Commercial Assays		
DyNAmo Flash SYBR Green qPCR Kit	Life Technologies	Cat# F-415L
High Pure Extender from Viral Nucleic Acid Large Volume Kit	Roche	Cat# 5114403001
MinElute PCR Purification Kit	QiaGEN	Cat# 28006
NextSeq 500/550 High Output Kit v2 (150 cycles)	Illumina	Cat# FC-404-2002
HiSeq 4000 SBS Kit (50/75 cycles)	Illumina	Cat# FC-410-1001/2
Deposited Data		
Raw and analyzed data (European Nucleotide Archive)	This study	ENA: PRJEB37490
Software and Algorithms		
AdapterRemoval v2	[88]	https://github.com/MikkelSchubert/adapterremoval
ADMIXTOOLS	[30]	https://github.com/DReichLab/AdmixTools
ANGSD	[32]	https://github.com/ANGSD/angsd
BWA	[89]	http://bio-bwa.sourceforge.net/
Dedup	[90]	https://eager.readthedocs.io/en/latest/
EAGER	[90]	https://eager.readthedocs.io/en/latest/
Geneious R8.1.974	[91]	https://www.geneious.com/
mapDamage2.0	[92]	https://ginolhac.github.io/mapDamage/
OptiType 1.3.1	[54]	https://github.com/FRED-2/OptiType
PMDtools	[93]	https://github.com/pontussk/PMDtools
Samtools	[94, 95]	http://samtools.sourceforge.net/
Schmutzi	[33]	https://grenaud.github.io/schmutzi/
SeqPrep	https://github.com/jstjohn/SeqPrep	https://github.com/jstjohn/SeqPrep
smartpca	[96]	https://www.hsph.harvard.edu/alkes-price/software/

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Johannes Krause (krause@shh.mpg.de).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

Data are available at the European Nucleotide Archive (ENA) under study accession number: PRJEB37490 (including the HBV genome recovered from individual SJN001 and the *Treponema pallidum* sub. *pertenue* genome recovered from individual SJN003).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The San José de los Naturales Royal Hospital (originally *Hospital Real de Sanct Joseph de los Naturales*) was founded between 1529 and 1531 [97–99] and dedicated to care exclusively for the indigenous population of the Viceroyalty of the New Spain. The establishment of this hospital was largely motivated by the need for medical facilities to aid the victims of the smallpox outbreaks that occurred early during the Colonial Period [100]. The vast majority of the patients were Nahuatl and Otomi speakers, which could tell us that most patients were from the central region of Mexico, including the valley of Morelos to the south and the Basin of the Balsas River to the west, as well as the boroughs of La Candelaria, Santo Tomás, San Pablo, San Antonio Abad and the towns of Jamaica and

Ixtacalco in Mexico City [22, 24]. The mortuary patterns of the archaeological context (multiple human skeletons in 13 out of 16 excavation units) suggests that most of the individuals probably died due to the epidemics that devastated Mexico City and the central plateau of New Spain which resulted in mass deaths [21, 101, 102].

Archaeological sites and sample description

Archaeological context

The samples used in this study were obtained from three individuals whose skeletons were recovered from the archaeological context of the San José de los Naturales Royal Hospital. These almost complete skeletons were recovered from the cemetery associated with the hospital during the excavations for a new subway line in downtown Mexico City between 1988 and 1994, under the supervision of the archaeologists Salvador Pulido Méndez and María de Jesús Sánchez Vázquez [21, 24]; during the season of 1992 the skeletons were found in the third stage of the excavation, suggesting they belonged to the oldest structures within the archaeological context.

Individuals and samples

Due to the possibility that these three individuals are of distinct ethnic origin (given that ethnic heterogeneity was promoted to generate social disintegration as a way to prevent the creation of communities and social organization among African slaves [2, 3, 5]). We collected one tooth from each of the following individuals: *ML8 SL 150* (age at death: 25-30 years old, male [25]; first right molar of the mandible, genomic library SJN001), *ML8 San José 214* (age at death: 30-35 years old, male (age at death: 30-35 years old, male [25]; second right molar of the maxilla, genomic library SJN002) and *ML8 SLU9B 296* (age at death: 25-30 years old, male [25]; second right molar of the mandible, genomic library SJN003). Samples were obtained under controlled conditions in the Osteology Laboratory of the Post Graduate Studies Division at the National School of Anthropology and History (ENAH) with a protocol devoted to minimize any possibility of contamination. Photographic records of the samples were obtained throughout the whole procedure. These individuals were previously described of potential African origin [6, 23, 25, 28] because of dental modifications (Figure 1) consistent with those practiced by sub-Saharan ethnic groups [25, 43, 103], as well as apparent dental and skeletal features with suggested affinity to sub Saharan human groups [26, 101]. The radiocarbon dates and strontium isotope analyses of the three individuals' samples were processed and analyzed at *The Curt-Engelhorn-Centre for Archaeometry* (Mannheim, Germany). Radiocarbon dates are reported showing the lab codes MAMS, and were processed using ultra filtrated collagen (fraction > 30kD) and dated using the *MICADAS-AMS* of the *Klaus-Tschira-Archäometrie Zentrum*. Strontium isotopes determination was performed in a high-resolution ICP mass spectrometer in a clean room laboratory at the same laboratory. Since this is quite a precise determination, as reflected by the reproducibility of NBS-987 measurements, we can confidently report and trust values to the fifth decimal place [63]. Osteological analyses were conducted on the skeletons in search for skeletal evidence of pathologies or trauma using the health indicators proposed by Goodman and Martin as well as references on osteological analyses of pathology [86, 104–106]. This methodology includes the identification of bone modifications in the skeleton present in the sites of muscle or ligament insertion caused by the hyperactivity of the main muscles responsible for the movement [107]. Dental modification patterns were assessed by comparing the dental decoration present in the SJN Africans with previously observed patterns [following the classic works by Reinaldo de Almeida (1953, 1957), as cited by references [41, 108]]. The individual biological affinity was estimated by cranial non-metric traits ancestry estimation, ranked by their degree of expression, with which percentages of biological affinity were obtained according to four biogeographical groups: Native American, African, Asian and European [109–111].

METHOD DETAILS

Osteological assessment

Osteobiography of individual ML8 SL 150 (SJN001)

Estimation of sex was performed by morphological analysis of sexually dimorphic features of the os coxa (pre-auricular sulcus, greater sciatic notch, subpubic morphology). Results indicate that this individual was likely a biologically male individual [112]. The age at death was estimated around 24 years and was obtained by analyzing the scalene tubercle and the first rib face (range of 24.28 to 51.86 years) [113]; the pubic symphysis (range: 22-25 years) [114] and the auricular surface of the ilium (range: 24-28 years) [115], which is consistent with previous estimates [25]. *Health:* The skull shows moderate cribra orbitalia and severe porotic hyperostosis. These abnormal skeletal changes are often in response to conditions associated with diet such as anemia and malnutrition, as well as parasitic infections and blood loss [38–40]. Observations investigating dental health reveal some signs of periodontal disease including receding of the alveoli with abnormal margins with associated alveolar inflammation. The tibiae and fibulae display dense sclerotic periosteal deposition suggestive of well-healed periosteal reaction. It is possible that these pathological changes are associated with conditions of poor hygiene, infectious processes, and malnutrition [40, 105, 107]. *Occupational activity:* The left clavicle presents enthesophyte development with associated raised margins at the insertion point of the coracoclavicular ligament, this may be attributed to remodelling of bone that occurs due to a microtraumas associated with repetitive use and has been reported to be found among individuals that they carry weighted loads on their shoulders [107]. The right clavicle presents exostosis in the anterolateral aspect located in the insertion of the coracoclavicular ligament (Figure S5A). The seventh thoracic vertebra presents Schmörl's node (Figure S5B) on the inferior aspect of the vertebral body resulting from the compression and subsequent hernia of inter-vertebral discs. Likewise, the ninth to the twelfth vertebrae present Schmörl's nodes on the superior and inferior aspects of the vertebral bodies. These nodes, or hernias, are the result of intense gradual compression and excessive

biomechanical stress [116]. *Cultural practices:* This individual shows dental modification in the four maxillary incisors (Figure 1 in the main text) made with two techniques: percussion fracture and filing, but also it can be seen some decoration work done in the mandibular canines [25, 107]. The patterns seen in the maxillary incisors are consistent with a "V" shaped modification present in Fang and other ethnic groups from the coastal region of central-Western Africa (present day Equatorial Guinea, northern Gabon, and southern Cameroon) [41, 43, 108]. The mandibular canines, however, present a dental modification pattern seen during 19th century expeditions in the Kingdom of Loango (present day Republic of the Congo), a multi-ethnic kingdom in the central-western part of Africa; and in Bakongo people from the Congo region [43]. *Taphonomy:* We observed green coloration on multiple skeletal elements acquired by contact with copper in the burial associated with the body. This green coloration was observed on the fifth and seventh cervical vertebra (Figure S5C) and the fourth and fifth left ribs (Figure S5D), there is a pattern of continuous inclined lines on the coloration. These indications are considered by other authors [23, 28] as a result of gun impacts. It is likely this individual had been shot and buried with the fragments in his body thus resulting in the green coloration of the skeleton. *Biological affinity:* The biological affinity of this individual was estimated from macromorphoscopic characteristics [109, 111], these characters are established by their degree of expression. For this work, the values obtained from the observation were analyzed using the software *Osteomics* [110], which bases its classification by a simple Bayesian method and proposes a probabilistic result taking into account the percentages of biological affinity according to four biogeographic groups: Native American, African, Asian and European. Individual 150 presented a large interorbital width (IOW); large nasal aperture width (NAW); as well as deep post-bregmatic depression (PBD); intermediate anterior nasal spine (ANS); inferior nasal aperture (INA) with border; a pronounced malar tubercle (MT); nasal bone contour (NBC) with intermediate plateau; supranasal suture (SPS) closed but visible; and zygomaticomaxillary suture (ZS) with an angle in the middle part; nasal overgrowth (NO) and transverse palatine suture (TPS) were not analyzed due to them being fractured. According to the classification analysis of *Osteomics*, a 92.23% probability of belonging to the African biogeographical group was found.

Osteobiography of individual ML8 San José 214 (SJN002)

The biological sex estimation of the individual was performed through observations of sexually dimorphic traits. Morphological analysis of the os coxa (preauricular sulcus, sciatic notch, composite arch, anterior edge form of the ischiopubic ramus and the ischiopubic ratio) [112] with particular attention to the morphology of the features associated with the pubic region (ventral arch, subpubic concavity, and the medial aspect of the ischiopubic ramus) resulted in the estimation of a male individual [117]. The age at death of the individual was estimated to be approximately 25 years and was calculated through age associated morphological observations of the scalene tubercle and the first rib face (range of 24.28 to 51.86 years) [113]; the pubic symphysis (range: 22-28 years) [114] and the auricular surface of the ilium (range: 25-29 years) [115]. *Health:* This individual presents signs of dental pathology including an abscess in the second lower right molar (Figure S6I), as well as periodontitis. The diaphyses of the lower limbs display periosteal deposition at the inferior most aspects. The right femora display sclerotic bone deposition creating an irregular surface which is suggestive of a healed periosteal reaction in life. Likewise, the tibiae and fibulae display periosteal reaction active at time of death; this is most apparent on the right tibia and fibula where the periosteal bone is deposited near and around the point of fracture with associated dense sclerotic and osteomyelitic bone development. *Degenerative diseases:* Degenerative disease of the joints is present and can be observed on the distal articular surface of the femur (Figure S6A) as osteochondritis dissecans, a defect of the articular surface of the bone due to compromise of the synovial capsule, and associated signs of osteoarthritis i.e., eburnation, lipping of the joint margins, and irregular joint surface [105]. Osteophytic development is present on the eighth, eleventh and twelfth thoracic vertebrae, and third and fourth lumbar vertebrae (Figure S6H); although these are age associated skeletal changes, development in younger individuals is associated with intense physical activity and excessive mechanical stress [107]. Here we see extreme osteophytic development and compression of the vertebrae. *Trauma:* Trauma refers to any sudden physical injury resulting from an external force. This individual, presents a healed fracture of the right tibia and fibula in the distal third of the diaphysis (Figures S6B–S6D) that extends from the lateral aspect to the medial aspect covered by and dense bony callus which is consistent with a healed oblique fracture and cross-section caused by the combination of axial compression force and angulation [107]. Healing of the fracture occurred without correct alignment of the fractured bones, which resulted in poor bone consolidation. The skull also presents a cut mark in the frontal bone with signs of bone regeneration (Figure S6J), which is indicative of the lesion happening ante mortem [118]. *Occupational activity:* The left and right clavicles display skeletal changes at the costoclavicular ligament, however, the right clavicle presents a higher degree of change with entheseal changes including depressed grained surface displaying raised irregular margins at the costoclavicular ligament (Figures S6E–S6G). The ulna presents enthesophyte development on the olecranon at the insertion of the triceps brachii, it is associated with continued biomechanical stress of the arm [107]. The lumbar vertebrae present Schmörl's nodes on the superior and inferior aspects of the vertebral bodies, resulting from the compression exerted on the inter-vertebral discs. In this individual, multiple Schmörl's nodes can be observed from the first to the fourth lumbar vertebrae. These vertebrae display skeletal changes (Figure S6H), as a result of an intense and gradual compression from biomechanical requirements including osteophytes on the superior and inferior surface and severe compression of the vertebra bodies in response to compression over time. The linea aspera of both femora show rugose enthesophyte development, a skeletal change associated with activity patterns; the significant enthesophytic development on the right femur is likely due to alterations in biomechanical activity resulting from the unaligned fracture of the proximal tibia and fibula. *Cultural practices:* This individual also presented dental modification in the four maxillary incisors (Figure 1 in the main text) from what appeared to be the fracture by percussion technique [107]. The patterns seen in the maxillary incisors are consistent with the "pattern No. 2" (resembling the "T" shape pattern proposed by Reichart et al. [108]) proposed by Wasterlain et al. [41], consisting in the removal of both mesial and distal incisal angles. This pattern was reported in D'zem (Njem) people, an ethnic group far to the east of the Dscha Valley (present day Cameroon) and their neighbors [43], but also

remarkably similar to that found in the individual no. 81 from the deposit of urban waste of the Valle da Gafaria (Lagos, Portugal) [41]. *Biological affinity:* The biological affinity of the individual was estimated from macromorphoscopic characteristics [109–111]. The individual shows a large interorbital width (IOW), a medium nasal aperture width (NAW), as well as deep post-bregmatic depression (PBD); small anterior nasal spine (ANS); inferior nasal aperture (INA) with a little border; a not so pronounced malar tubercle (MT); a rounded nasal bone contour (NBC); a closed but visible supranasal suture (SPS); and zygomaticomaxillary suture (ZS) with an angle in the middle part; nasal overgrowth (NO) and transverse palatine suture (TPS) were not analyzed due to them being fractured [109, 111]. According to the classification analysis of *Osteomics*, a 95.24% probability of belonging to the African biogeographical group was found.

Osteobiography of individual ML8 SLU9B 296 (SJN003)

The sex estimation of individual SJN003 was performed through morphological observations of the os coxa (preauricular sulcus, sciatic notch, composite arch, anterior edge form of the ischiopubic ramus and the ischiopubic ratio) [112] including the morphology of the pubic region (ventral arch, subpubic concavity, and the medial aspect of the ischiopubic ramus) [117]. All observations resulted in estimations of a biologically male individual. The age at death was estimated to be approximately 25.26 years and was calculated by analyzing the scalene tubercle and the first rib face [113], which is consistent with previous estimates [25]. It was not possible to obtain age using another method due to preservation issues of the pubic symphysis and the auricular surface of the ilium. *Health:* The skull presents moderate cribra orbitalia. The number of permanent teeth observed is 13 and an alveolus; abscesses or injured teeth were not present and periodontitis is observed. The presence of cribra orbitalia and hyperostosis in the skeleton can be due to three factors: diet such as malnutrition and anemia, parasitic infections, and blood loss [86]. The diaphyses of the long bones present slight infectious processes, with the exception of the femurs in which osteomyelitis can be observed on the anterior and posterior sides of the shaft (Figure S7A). Periosteal changes could be associated with a variety of physiological and environmental stressors including diet or infectious processes [86]. *Occupational activity:* The left clavicle diaphysis (Figure S7B) presents diffuse cortical irregularity and a roughened, raised appearance at the muscular insertion of the deltoid [86]. *Cultural practices:* This individual presents dental modification of maxillary and mandibular canines and central and lateral incisors (Figure 1 in the main text). The patterns seen resemble the ‘‘pattern No. 2’’ (resembling the ‘‘T’’ shape pattern proposed by Reichart et al. [108]) proposed by Wasterlain et al. [41], consisting in the removal of both mesial and distal incisal angles. This pattern was reported in D’zem (Njem) people, an ethnic group far to the east of the Dscha Valley (present day Cameroon) and their neighbors [43], and very similar to that found in the individual no. 81 from the deposit of urban waste of the Valle da Gafaria (Lagos, Portugal) [41]. *Biological affinity:* The biological affinity of the individual was estimated from macromorphoscopic characteristics [109–111]. The individual shows a large interorbital width (IOW), a medium nasal aperture width (NAW), as well as deep post-bregmatic depression (PBD); medium anterior nasal spine (ANS); inferior nasal aperture (INA) with a pronounced border; a not so pronounced malar tubercle (MT); a slightly rounded nasal bone contour (NBC) with high walls; a closed but visible supranasal suture (SPS); transverse palatine suture (TPS) goes all way through and reaches the medium line, where it projects to both the anterior and the posterior ends drawing a M shaped line; and zygomaticomaxillary suture (ZS) with an angle in the middle part; nasal overgrowth (NO) was not analyzed due to it being fractured [109, 111]. According to the classification analysis of *Osteomics*, a 91.95% probability of belonging to the African biogeographical group was found.

Stable isotope analysis, diets, and the reservoir effect

Stable isotope analysis of human bone can provide insights into the significance of different resources to the protein component of the diet, this includes the potential reliance on marine resources and their impact on ^{14}C radiocarbon measurements through the ‘reservoir effect’ [119, 120]. Here, we analyzed dentine collagen from the same teeth used to conduct the radiocarbon dating and aDNA analysis. These teeth were first (SJN001) and second molars (SJN002 and SJN003), providing dietary information for the years of tooth formation (crown-enamel formation at 2-2.5 years old for the first molar, and between 7-8 years old for the second molar) [121, 122]. Sample SJN001, as a first molar, should be treated with caution as high $\delta^{15}N$ values could be the product of breastfeeding which results in a higher trophic level appearance [34]. We obtained 250 mg of dentine powder from the crown of the analyzed teeth using a diamond-tipped drill. Collagen was then extracted from this powder using standard procedures [123]. The powder was demineralised in 10 mL aliquots of 0.5M HCl at 4°C. The acid was changed until CO_2 stopped evolving. The residue was rinsed three times in deionised water before being gelatinized in pH = 3 HCl at 75°C for 48 hours. The resulting solution was filtered, with the supernatant then being lyophilized over a period of 24 hours. After calculating the collagen yield, all purified collagen samples (1 mg) were located in tin capsules to be analyzed in duplicate at the Department of Archaeology, Max Planck Institute for the Science of Human History using a Thermo Fisher™ Elemental Analyzer coupled to a Thermo Fisher™ Delta V™ Advantage Mass Spectrometer via a ConFloIV system. Accuracy was determined by measurements of international standard reference materials within each analytical run. These were USGS40 $^{13}C_{raw} = -26.4 \pm 0.1$, $^{13}C_{true} = -26.4 \pm 0.0$, $^{15}N_{raw} = -4.4 \pm 0.1$, $^{15}N_{true} = -4.5 \pm 0.2$; IAEA N2 $^{15}N_{raw} = 20.2 \pm 0.1$, $^{15}N_{true} = 20.3 \pm 0.2$; IAEA C6 $^{13}C_{raw} = -10.9 \pm 0.1$, $^{13}C_{true} = -10.8 \pm 0.0$. The atomic C:N ratio along with the collagen yields were used in order to determine the quality of collagen preservation. Collagen yields over 1 wt% are considered acceptable for carbon and nitrogen values [124], while the C:N ratio should range from 2.9 to 3.6 [125]. Stable carbon and nitrogen isotope ratios of human tissues are expressed in δ notation relative to established international standards and shown in parts per thousand (‰). The delta values are obtained by the following equation:

$$\delta = \frac{R_{sample} - R_{standard}}{R_{standard}} \times 1000$$

where the R represents the ratio between the heavier (^{13}C, ^{15}N) and lighter (^{12}C, ^{14}N) isotopes [126]. $\delta^{13}C$ and $\delta^{15}N$ analysis of human bone collagen primarily reflects the isotopic values of the protein input to the diet, with minimal contribution of lipids and carbohydrates, meaning that it will be heavily influenced by protein-rich foods [127].

Stable carbon isotope variability in terrestrial ecosystems is primarily driven by two dominant photosynthetic pathways, C_3 and C_4, which differ in their net discrimination against ^{13}C during CO_2 fixation [128]. In C_3 plants, strong discrimination against ^{13}C results in lower ^{13}C values in virtually all trees, shrubs, and temperate grasses, including wheat, than in C_4 plants such as maize and millet [129]. C_3 ^{13}C values vary from *c.* −24 to −36‰ (global mean −26.5‰), while C_4 values range from *c.* −9 to −17‰ (global mean −12‰) [129]. C_3 and C_4 plants thus have distinct and non-overlapping ^{13}C values [130]. These distinctions are reflected in the tissues of consumers, with small trophic level effects of 1-2‰ [127].

Stable nitrogen isotope ratios vary with tropic level, and ^{15}N trophic shifts of +2-6‰ from plants to herbivores, and from herbivores to carnivores, is well documented in marine and terrestrial systems [35, 131]. This trophic effect is most likely linked to the loss of ^{15}N-depleted excretion products [132], although diet-tissue distinctions are highly variable between animals [133]. The long length of marine food chains, leads to distinctively high ^{15}N in marine foods and consumers compared to their terrestrial counterparts [134]. Freshwater foods also tend to have high ^{15}N though ^{13}C does not follow the same trend toward higher measurements as in marine food chains due to different sources of carbon dioxide for primary producers [135].

Despite the long history of stable isotope research in African archaeology, there is a significant inequality in the application of these studies across the regions of Africa [36]. While South and East Africa have been well covered, other regions, especially West Africa, remain relatively unexplored [36]. In the context of 16th-19th century slavery, historical records are often lacking in terms of insights into the origins of enslaved individuals and their subsistence strategies, focusing on coastal embarkation points into the Atlantic Trade rather than places of residence which may, in many cases, have been hundreds of kilometres away [136]. Stable carbon and nitrogen isotope analysis of 16th-19th century slave populations in southern Africa [137] and the Caribbean [67] have already shown the efficacy of this methodology for uncovering insights into individual origins.

In the context of 16th-19th West Africa, stable isotopic insights into broad subsistence bases may also provide geographic information. For example, along the west coast there was a well-documented focus on rice (C_3) [138], in the south of Ghana there was a greater focus on yams and other root crops (C_3), and in the semi-arid interior there was often a contribution of C_4 crops such as sorghum and millet [139]. Thus, the different regional agriculture techniques of subsistence may be useful in establishing the origin of those individuals enslaved in the West of Africa [67]. That said, existing isotopic information from slaves of African origin in the Americas indicate an adult diet of C_4 crops and marine resources [64, 67, 140, 141]. This is particularly significant in our context given its potential influence on establishing a secure chronology for human remains.

Marine and freshwater environments are reservoirs of ^{14}C. This means that in a marine or freshwater environment, ^{14}C concentrations in CO_2 are lower than contemporary atmospheric values, leading to 'older' radiocarbon dates in those organisms relying on freshwater or marine systems [142]. Aquatic plants and phytoplankton fix this CO_2, transferring relatively low concentrations of ^{14}C up the foodchain. The direct or indirect consumption of marine proteins by human beings will thus impact ^{14}C measurements of their tissues, making them seem 'older' through the marine reservoir effect [143–145]. Thus, before the conversion of the radiocarbon date results to a calendar age, it is important to determine the degree of marine consumption in an individual.

Ancient DNA sample processing and quality control

Sampling

All samples were processed in dedicated laboratories at the Max Planck Institute for the Science of Human History in Jena, Germany. Tooth powder (~80 mg) was obtained by cutting each dental piece at the junction between the root and crown and sampling the dental pulp from the crown (*A libraries*) and the root (*B libraries*). These procedures were carried out after a bleach/rinse and UV decontamination protocol. After sampling, the remaining root was used for radiocarbon dating, while the crown was sampled for Sr determinations as previously described.

DNA extraction

DNA was obtained from bone powder after overnight extraction at 37°C with a decalcifying buffer (900 μL EDTA 0.5M, 75μL H_2O and 25 μL Proteinase K) [146]. The whole mixture was then centrifuged to pellet the remaining tooth powder and the liquid transferred into a binding buffer as previously described [147]. DNA was then purified by a silica column-based method using silica columns for high volumes (*High Pure Viral Nucleic Acid Large Volume Kit*, Roche Molecular Systems, Inc.; Pleasanton, CA). DNA was eluted in TET (10mM Tris, 1mM EDTA and 0.05% Tween) in two 50 μL-steps for a final volume of 100 μL and frozen at −20°C until library preparation. Negative blanks and cave bear positive controls were included for each step of the procedure.

Library preparation

We built one non-uracil DNA glycosylase (non-UDG) treated library using 15 μL of each DNA extract to assess the authenticity of the extracted DNA after obtaining the characteristic damage plots associated with ancient DNA [148]. We then used 20 μL of each DNA extract to build UDG-half libraries with *Illumina*-specific adapters following a modified double-stranded library preparation protocol as previously described [149, 150]. Each library was treated independently despite being from the same individual. Libraries were quantified using qPCR with the IS7 and IS8 primers in a quantification assay using a *DyNAmo SYBR Green qPCR Kit* (Thermo Fisher

Scientific Inc.; Waltham, MA) on the *LightCycler® 96* (Roche Diagnostics Inc.; Risch-Rotkreuz, Switzerland). Each library was identified with the respective pair of indexes in double-100 μL reactions using *PfuTurbo DNA Polymerase* (Agilent Technologies, Inc.; Santa Clara, CA). The indexed products for each library were pooled, purified over silica columns using the *MinElute PCR Purification Kit* (QIAGEN N.V.; Hilden, Germany), eluted in 44 μL TET and again qPCR quantified, now using the IS5 and IS6 primers. Conditioning for sequencing included the amplification of the purified product in 4x100 μL reactions using *Herculase II Fusion DNA Polymerase* (Agilent Technologies, Inc.; Santa Clara, CA) following the manufacturer's specifications with 0.3 μM of each IS5/IS6 primers, following a purification over silica columns also using the *MinElute PCR Purification Kit*, and elution in a final volume of 22 μL TET. Two microliters of the conditioned product were diluted 1:10 and quantified using the *Agilent 2100 Bioanalyzer DNA 1000* protocol (Agilent Technologies, Inc.; Santa Clara, CA). An equimolar (10mM final concentration) pool of all libraries was then prepared for shotgun sequencing on the *Illumina HiSeq 4000 Systems* platform (Illumina, Inc., San Diego, CA). Both libraries of each individual were sequenced to 5 million reads depth to obtain basic QC parameters with the aid of the *EAGER* pipeline ver. 1.92.55 [90] and decided for either deeper sequencing or in-solution capture.

Whole-genome and immune-genes captures

Using in-solution capture based on modified immortalized probe sequences [151], target immunity genes sequences or a panel of around 1,237,207 single nucleotide polymorphisms (SNPs) were enriched via in-solution capture from the total DNA in the sequencing libraries [152–154]. Briefly, UDG-half- treated libraries were re-conditioned by further amplifying with IS5/IS6 primers and *Herculase II Fusion DNA Polymerase* to reach a concentration of 200–400 ng/μL as measured on a *NanoDrop 8000* spectrophotometer (Thermo Fisher Scientific Inc.; Waltham, MA). Capture was performed on 5.25 μL of each re conditioned library using that volume for each capture procedure [153, 154]. After enrichment, captured library pools were paired-end sequenced on the *Illumina Hiseq 4000* (Illumina, Inc., San Diego, CA) with 75 single-end cycles providing on average 10 million reads per sample. *Illumina* sequencing adapters were removed by demultiplexing, which was performed by sorting all the sequences corresponding to their respective indices using the *bcl2fastq Conversion Software* ver. 2.17.1.14 and *dnaclust* ver. 3.0.0 [155].

Immune-genes capture design

We designed a set of enrichment probes for 488 human genes part of both the innate and adaptive immune system [156]. Exon sequences for these genes were extracted from the human genome build *hg19* [29] using the *RefSeqGene* records from the NCBI/Nucleotide database [157]. Introns were also used in the case of MIC genes. Given the polymorphic nature of these genes, we added alternative alleles for HLA, MIC, TAP and KIR genomic regions, which were obtained from the IMGT/HLA database [158]. For HLA class I and KIR genes the intronic regions were also included. For the HLA and MIC genes a set of 83 representative alleles with full-length gene sequences was chosen that encompasses the major allelic groups [159] and covers 95% of the known polymorphism. To capture the remaining 5%, a set of 162 × 160 base pairs (bp) consensus sequences was designed. A 60-bp probe set was designed at every 5bp interval along the target sequences. The final 52-bp probe sequences were mapped to *hg19* using RazerS3 [160] with a minimum threshold of 95% identity. The probe set was tripled to complete capacity of the one-million feature SureSelect DNA Capture Array, which was turned into an in-solution DNA capture library as described elsewhere [153]. We enriched the reads mapping to the HLA region with the described in-solution capture approach, yielding 10X-100X coverage for this genomic region. For dealing with the allele assignation coming from short reads, we used Optitype [54], a software specifically developed to deal with short reads coming from ancient DNA data. Our capture approach was validated using previously typed samples coming from an external quality control program (UCLA's International HLA DNA Exchange), for which the DNA was shredded using the Covaris system to provide short DNA fragments for a proper validation of the results. The fact that the alleles typed in our individuals are commonly found in Sub-Saharan African populations further supports our findings.

Treponema pallidum capture design

For targeted enrichment of *Treponema pallidum* DNA probes were designed on the basis of *Treponema pallidum* subsp. *pallidum*, *T. pallidum* subsp. *endemicum* and *T. pallidum* subsp. *pertenue* strains. The probes were designed with a 1bp tiling and a length of 52 bp with an additional 8bp linker sequence (CACTGCGG) as described previously [153]. Duplicated probes and probes with low sequence complexity were removed. This resulted in 1,125,985 unique probe sequences. This probe set was spread on two Agilent one-million feature SureSelect DNA Capture Arrays. The capacity of the two arrays was filled by randomly duplicating probes from the probe set. The arrays were turned into an in-solution DNA capture library as described elsewhere [153].

List of Treponema pallidum genomes IDs used for probe design

Treponema pallidum subsp. *pallidum* strains Nichols (NC_000919.1), SS14 (NC_021508.1), Sea 81-4 (NZ_CP003679.1), Mexico A (NC_018722.1), *T. pallidum* subsp. *endemicum* strain Bosnia A (NZ_CP007548.1), and *T. pallidum* subsp. *pertenue* strain Fribourg-Blanc (NC_021179.1).

Hepatitis B Virus (HBV) capture design

For targeted HBV DNA enrichment, probes were designed on the basis of a worldwide set of HBV full genomes (see list of GenBank accessions below). The probes were designed with a 1bp tiling and a length of 52 bp with an additional 8bp linker sequence (CACTGCGG) as described previously [153]. Duplicated probes and probes with low sequence complexity were removed. This resulted in 221,190 unique probe sequences. This probe set was replicated four times on an Agilent one-million feature SureSelect DNA Capture Array, which was turned into an in-solution DNA capture library as described elsewhere [153].

List of HBV genome IDs used for probe design

Petersberg, Sorsum, Karsdorf, LT992459, LT992455, LT992454, LT992448, LT992447, LT992444, LT992443, LT992442, LT992441, LT992440, LT992439, LT992438, JN315779, MG585269, AB076679, AB116084, AB453988, AY738142, GQ477499, AY934764,

FJ692556, FJ692598, FJ692611, GQ161813, GQ331046, AB073858, AB033555, AB219429, AB219430, AP011089, AB073835, AB287316, AB287318, AB287320, AB287321, DQ463789, DQ463792, AB241117, DQ993686, AB111946, AB112066, AB112472, DQ089767, X75656, X75665, AB048704, AB048705, AF241411, AP011100, AP011102, AP011103, AP011106, AP011108, FJ899792, JN642140, GQ477453, GQ477455, JN642160, JN642163, JN688710, JN688711, GQ922005, HE974378, KJ470893, KJ470896, KJ470898, FJ904430, FJ904436, AB033559, AB048701, AB048702, AB188243, AB210818, AM494716, AY796031, AY902768, DQ315779, X80925, X75657, X75664, AY090458, AB116654, FJ657525, AY090455, AY311369, DQ899144, DQ899146, AB116549, X75663, AF223962, AB166850, AB056513, AB064312, AF405706, AB059660, AB375163, AY090454, AY090457, AB486012, AY330911, AJ131571, AY781180, U46935, AJ131567, AF193863, EU155824, KC790378, KC790377, KC790376, NC_001484, U29144, NC_004107, AF046996, KY703886, MH307930, AF498266.

Pathogen genome capture

After finding reads mapping to HBV (SJN001) and *Treponema pallidum* (SJN003), we selectively enriched for these pathogens in their respective libraries. Using in-solution capture based on modified immortalized probe sequences [153], target pathogen genomes were enriched via in-solution capture from the total DNA in the sequencing libraries [152–154]. Briefly, UDG-half- treated libraries were reconditioned by further amplifying with IS5/IS6 primers and *Herculase II Fusion DNA Polymerase* to reach a concentration of 200–400 ng/μL as measured on a *NanoDrop 8000* spectrophotometer (Thermo Fisher Scientific Inc.; Waltham, MA). Capture was performed on 5.25 μL of each re-conditioned library using that volume for each capture procedure [153, 154]. For targeted HBV and *Treponema* genome DNA enrichment, we used the previously described sets of probes. After enrichment, captured library pools were paired-end sequenced on the *Illumina Hiseq 4000* (Illumina, Inc., San Diego, CA) with 75 cycles providing on average 10 million reads per sample for the first round of capture and 20 million (*Treponema*) and 13 million (HBV) reads per sample for the second round of capture.

QUANTIFICATION AND STATISTICAL ANALYSIS

QC and data processing for human PopGen

We performed an analysis of the human captured sequence data using *EAGER* ver. 1.92.55 [90]. Clip&Merge [90] and *AdapterRemoval* v.2 [88] were used to trim adaptor sequences and to remove adaptor dimers and low-quality sequence reads (min length = 30; min base quality = 20) from the reads resulting from the sequencing of the aDNA libraries. Pre-processed sequences were mapped to the human genome assembly GRCh37 (hg19) from the Genome Reference Consortium [29] using *BWA* ver. 0.7.12 [89] and a default seed length of 32.

aDNA Authentication

The C to T misincorporation frequencies typical of aDNA were obtained using *mapDamage* 2.0 [92] to assess the authenticity of the ancient DNA fragments from the non-UDG treated libraries. Characteristic short average fragment length (49-55 base pairs) and the increased proportion of miscoding lesions due to deamination at the molecule extremes were found on the six UDG-half libraries analyzed (Table S2).

Data processing

Duplicates were removed with *DeDup* [90], which removes identical reads. UDG-half libraries were trimmed for the first and last three positions to reduce the impact of deamination-induced misincorporations during genotyping. We used *samtools mpileup* (parameters –q 30 –Q 30 –B) to generate a pileup file from the merged sequence data of each individual, and used a custom script [*pileupCaller* ver. 8.2.2 [161]] to genotype the individuals, using a pseudo-haploid random draw approach. For each position on our capture panel, a random read was drawn for each individual and the allele of that read was assumed to be the homozygous genotype of the individual at that position. In order to compare with available data from African populations, were merged our SNPs to the approx. 600,000 SNPs of the Human Origins dataset [30]. Genetic sex of the three samples was assigned using SNP capture data by calculating the ratio of average X chromosomal and Y chromosomal coverage to average autosomal coverage at the targeted SNPs [31]. Samples with an X rate between 0.35 and 0.55 and a Y rate between 0.4 and 0.7 were confirmed male. Since the three analyzed individuals were male, *ANGSD* was run to measure the rate of heterozygosity of polymorphic sites on the X chromosome after accounting for sequencing errors in the flanking regions to estimate the nuclear contamination, since males are expected to have only one allele at each position [32]. Reads mapping to the human mitochondrial DNA (mtDNA) were used to reconstruct a mtDNA consensus sequence and estimate contamination levels with *schmutzi* [33]. Present-day human contamination estimates were performed using a comparative database of 197 modern-day worldwide mtDNA sequences provided with the software package. After assessing the quality of the independent libraries, libraries A and B of each individual were merged using *SAMtools* ver. 1.3 [94].

Uniparental markers

Mitochondrial DNA (mtDNA) genomes were determined mapping reads to the revised Cambridge reference sequence [162]. For the resulting sequences, we filtered positions with likelihoods above 10, 20 or 30 and used *HaploGrep2* [163] and *HAPLOFIND* [164] to assign and confirm the corresponding mtDNA haplogroups. The aligned mtDNA genomes were compared to other available genomes from literature which belong to one of the three haplogroups retrieved (153 sequences for L1b, 208 for L3d and 415 for L3e). The mtDNA sequences were aligned with MAFFT (http://mafft.cbrc.jp/alignment/software/) [165] and manually checked with Bioedit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). The two poly-C regions (positions: 303–315, 16,183–16,194) were excluded from the analysis. Haplotype based networks (Figure S1) were built with R with the command haploNet of package PEGAS [166]. Sequenced reads mapped to Y-Chromosome (Y-Chr) SNPs within our capture panel that are also present in the Y-DNA

Haplogroup Tree 2018 ver. 13.174 [167] were used to assign Y-Chr haplogroups for each individual. Haplogroup assignment was manually confirmed by looking at the most downstream SNP retrieved after assessing the presence of upstream mutations along the Y-Chr haplogroup phylogeny.

Principal components analysis

smartpca (version 16000) from the *Eigensoft* package [168] was used to calculate Principal Components (PCs) of variation in the present-day populations from the HO dataset, using the options "lsqproject: YES" and "shrinkmode: YES." Initially, we projected the three ancient individuals on PCs calculated on the genetic variation in 371 worldwide populations, to access the continental-level ancestries in the ancient individuals (Figure S2). We then projected the ancient individuals on PCs calculated on variation from 534 Africans from 51 populations, both from Northern and sub-Saharan Africa (Figures 2 and S2).

Worldwide populations

Abazin, Abkhasian, Adygei, Afar WGA, Ain Touta WGA, Albanian, Aleut, Aleut Tlingit, Algerian, Algonquin, Altaian, Altaian Chelkans, Ami, Apalai, Arara, Argentina Puna, Armenian, Armenian Hemsheni, Assyrian, Atayal, Australian WGA, Australian, Avar, Aymara, Azeri, Azeri WGA, Baalberge MN, Balkar, Balochi, Bantu Herero, Bantu Kenya, Bantu SA, Bantu SA Herero, Bantu SA Ovambo, Bantu Tswana, Bashkir, Basque, Bedouin A, Bedouin B, Belarusian, Bengali, Bergamo, Berry Au Bac, Besermyan, Biaka, Bichon, Bockstein published, Bolivian, Borneo, Bougainville, Brahmin Tiwari, Brahmin, Brahui, Brillenhohle, Bulgarian, Burbur WGA, Burmese, Burusho, Buryat, Cabecar, Cambodian, Canary Islander, Chane, Chaudardes1, Chechen, Chilote, Chipewyan, Choiseul, Chukchi, Chuvash, Circassian, Cree, Crete, Croatian, Cypriot, Czech, Dai, Damara, Darginian, Datog, Daur, Dinka, Dolgan, Druze, Dungan, Dusun, Egyptian, Enets, English, Esan, Eskimo Chaplin, Eskimo Chaplin Sireniki, Eskimo Naukan, Eskimo Sireniki, Estonian, Even, Evenk FarEast, Evenk Transbaikal, Ezid, Falkenstein, Finnish, French, Fuego Patagonian.SG, Gagauz, Gambian, Gana, Georgian, Georgian WGA, Georgian, German, Greek, Greek WGA, Guarani, Guarani GN, Guarani KW, Gui, Gujarati A, Gujarati B, Gujarati C, Gujarati D, Hadza, Hadza1, Haiom, Han, Hawaiian, Hazara, Hezhen, Himba, Hoan, Hungarian, Icelandic, Igbo, Igorot, Inga, Ingushian, Iran contemporary, Iran Zoroastrian, Iranian, Iraqi Jew, Irish, Irish Ulster, Irula, Italian North, Italian South, Itelmen, Japanese, Jew Ashkenazi, Jew Cochin, Jew Ethiopian, Jew Georgian, Jew Iranian, Jew Iraqi, Jew Libyan, Jew Moroccan, Jew Tunisian, Jew Turkish, Jew Yemenite, Jordanian, Ju Hoan North, Ju Hoan South, Kabardinian, Kaitag, Kalash, Kalmyk, Kapu, Kaqchikel, Karachai, Karaim, Karakalpak, Karelian, Karitiana, Kazakh, Ket, Kgalagadi, Khakass, Khakass Kachins, Khamnegan, Kharia, Khomani, Khwe, Kikuyu, Kinh, Kirghiz, Kolombangara, Kongo, Korean, Koryak, Kotias, Kubachinian, Kumyk, Kurd, Kurumba, Kusunda, Kyrgyz, Lahu, Lak, Lapita Tonga, Lebanese, Lebanese Muslim, Lemande, Lezgin, Libyan, Lithuanian, Lodhi, Loschbour, Luhya, Luo, Makira, Makrani, Mala, Malaita, Maltese, Mamanwa, Mandenka, Mansi, Maori, Masai, Mayan, Mbuti, Mende, Miao, Mixe, Mixtec, Moldavian, Mongol, Mongola, Mordovian, Moroccan, Mozabite, Mycenaean, Nahua, Nama, Nanai, Naro, Nasioi, Naxi, Negidal, New Guinea, Nganasan, Nggela, Nivh, Nogai, Norwegian, Oase1, Ojibwa, Okunevo.SG, Onge, Ontong Java, Orcadian, Oromo, Oroqen, Ossetian, Ostuni1, Palestinian, Papuan, Papuan Central, Papuan Gulf, Pathan, Piapoco, Pima, Polish, Poltavka, Potapovka, PPNB, PPNC, Punjabi, Quechua, Ranongga, Relli, Remedello BA, Rennell and Bellona, Rochedane published, Romanian, Russell, Russian, Russian Archangelsk Krasnoborsky, Russian Archangelsk Leshukonsky, Russian Archangelsk Pinezhsky, Saami WGA, Saharawi, Samaritan, Sandawe, Santa Cruz, Santa Isabel, Sardinian, Satsurblia, Saudi, Savo, Saxon, Scottish, Scythian IA, Selkup, Semende, Shaigi WGA, She, Sherpa, Shetlandic, Shor Khakassia, Shor Mountain, Shua, Sicilian, Sindhi Pakistan, Somali, Sorb, Spanish, Spanish North, Srubnaya, Stuttgart, Surui, Sweden IA.SG, Syrian, Taa East, Taa North, Taa West, Tabasaran, Tajik, Tatar Astrakhan, Tatar Crimean, Tatar Kazan, Tatar Mishar, Tatar Siberian, Tatar Siberian Zabolotniye, Thai, Tibetan, Ticuna, Tikopia, Tlingit, Todzin, Tongan, TRB Sweden MN.SG, Tshwa, Tswana, Tu, Tubalar, Tujia, Tunisian, Turkish, Turkish Balikesir, Turkmen, Tuscan, Tuvinian, Udmurt, Ukrainian, Ulchi, Urubu Kaapor, Uygur, Uzbek, Uzbek WGA, Vatya.SG, Vella Lavella, Veps, Vestonice13, Vishwabrahmin, Wambo, Wayuu, Xavante, Xibo, Xibo, Xuun, Yadava, Yakut, Yemeni, Yemenite Jew, Yi, Yoruba, Yukagir, Zapotec, Zoro.

African and related populations

Algerian, Bantu (Kenya), Bantu (South Africa), Bantu (Herero), Bantu (Ovambo), Bedouin A, Bedouin B, Biaka, Canary Islander, Damara, Datog, Dinka, Egyptian, Esan, Gambian, Gana, Gui, Hadza, Haiom, Himba, Hoan, Igbo, Ethiopian Jew, Libyan Jew, Moroccan Jew, Tunisian Jew, Ju hoan (North), Ju hoan (South), Kgalagadi, Khomani, Khwe, Kikuyu, Kongo, Lemande, Libyan, Luhya, Luo, Mandenka, Masai, Mbuti, Mende, Moroccan, Mozabite, Nama, Naro, Oromo, Saharawi, Sandawe, Shua, SJN, Somali, Taa East, Taa North, Taa West, Tshwa, Tswana, Tunisian, Wambo, Xuun, Yoruba.

ADMIXTURE analysis

We used ADMIXTURE ver. 1.3.0 [169], a maximum-likelihood based clustering algorithm to estimate the genetic structure present in our samples, after excluding variants with minor allele frequency of 0.01 and following LD pruning using Plink (ver. 1.90) with a step size of 5, a window size of 200, and an R2 threshold of 0.5 [170]. For K = 2 to K = 14, we estimated the cross-validation (CV) error with 100 bootstrap replicates in an unsupervised model using a panel of 66 populations, including two European (Spanish and French), two Asian (Han and Cambodian), one Oceanian (New Guinea) and two Native American (Mixe and Zapotec) populations to account for different probable sources of genetic contribution in the context of the colonial period in Mexico [2, 5, 6, 26, 101, 168, 171] to assess whether the analyzed individuals had traces of admixture from non-African parental populations (Figure 2). The lowest CV error corresponded to K = 10.

Population dataset for ADMIXTURE

Algerian, Bantu (Kenya), Bantu (South Africa), Bantu (Herero), Bantu (Ovambo), Bedouin A, Bedouin B, Biaka, Cambodian, Canary Islander, Damara, Datog, Dinka, Egyptian, Esan, French, Gambian, Gana, Gui, Hadza, Haiom, Han, Himba, Hoan, Igbo, Ethiopian Jew, Libyan Jew, Moroccan Jew, Tunisian Jew, Ju hoan (North), Ju hoan (South), Kgalagadi, Khomani, Khwe, Kikuyu, Kongo,

Lemande, Libyan, Luhya, Luo, Mandenka, Masai, Mbuti, Mende, Mixe, Moroccan, Mozabite, Nama, Naro, New Guinea, Oromo, Saharawi, Sandawe, Shua, SJN, Somali, Spanish, Taa East, Taa North, Taa West, Tshwa, Tswana, Tunisian, Wambo, Xuun, Yoruba, Zapotec.

F_3 and D tests

To assess the genetic relationships and admixtures suggested in the PCA and ADMIXTURE analysis, we carried out *F*-statistics using the programs qp3Pop ver. 412 and qpDstat ver. 711 in the ADMIXTOOLS suite [172] for F_3- and D-statistics, respectively. We tested each individual independently, as they most probably were of different ancestries as explained before. We used a F_3-statistics of the form f_3 (*Outgroup*; *X*,*Y*) to measure the amount of shared genetic drift of populations *X* and *Y* after their divergence from a non-African outgroup (Figure 3); were *X* is each of the SJN individuals, *Y* is each of the populations in our database, and *Outgroup* is the genome of the Ust'-Ishim individual [62]. D-statistics of the form *D* (*Chimp*, SJN00*X*; *Y*, *Target*) were used to demonstrate if our African individuals are related to a target population or shared an excess of alleles with any population in position *Y*. A negative value implies that either *Chimp* and *Target*, or *SJN* and *Y* share more alleles than expected under the null hypothesis of a symmetrical relationship between *Y* and *Target*. Target was each of the populations with the highest top F_3 values for each SJN individual. For SJN001 (Figure S3A) Mende were used as Target; for SJN002 (Figure S3B), Bantu-speaking Ovambo were used as Target; for SJN003 (Figure S3C) Lemande were used as Target. To test for cladality between each of the individuals from SJN and their respective Target populations, we computed *D* statistics of the form *D* (*Chimp*, *X*; *Target*, *SJN00X*). We find that none of the ancient individuals are truly cladal with the *Target* populations (Table S1).

Pathogen genomes assemblies

Hepatitis B Virus genome

After HBV DNA capture and sequencing, resulting reads were merged and aligned against an HBV genome (GenBank accession: KC875253) using Geneious v. 9 with medium sensitivity settings. A 90% consensus sequence was generated with a coverage threshold of 3x. Reads were remapped against the consensus sequence within the Eager pipeline [90] to perform final coverage and damage assessment. The sequence was aligned with a worldwide set of HBV full genomes (see accession numbers below), using MAFFT [165]. The alignment was cleaned using Gblocks [173], removing positions with more than 50% gaps. A phylogenetic tree was constructed using RAxML v. 8 [174] with GTRCAT substitution model and the rapid bootstrap algorithm (Figure 4).

Treponema pallidum sub. pertenue genome and phylogenetic analyses

Shotgun reads sequenced from both libraries of SJN003 were screened for pathogens in MALT [175]. Reads mapping to *Treponema pallidum* were identified and their ancient origin was validated from deamination of cytosine to thymine bases at the reads' ends. A whole genome-wide capture with the probes described above was performed to enrich for DNA fragments of treponemal species. Paired-end reads from reference genomes; ancient genomes reads from individuals 94B and 133 from the ex-Convent of Santa Isabel from colonial Mexico [71] and reads from second round of enrichment of the SJN003 libraries were processed in EAGER (version 1.92.7) [90]. Additionally, simulated reads (100bp length with 99bp overlap to successive reads) from complete genomes of *Treponema pallidum* downloaded from NCBI were analyzed. Adapters and poor quality (minimum base quality of 20 and sequence length of 30) reads were filtered out. Both simulated and sequenced reads were mapped against *Treponema pallidum* sub. *pallidum* Nichols reference genome (GenBank accession number: NC_021490.2) using BWA ver. 0.7.12 [89] for a seed length of 32, a 0.1 mapping stringency setting (-n) and mapping quality of 37 followed by exclusion of duplicated reads using Mark Duplicates [176]. GATK identified genotypes for all sites and variants were estimated. Variants from all *vcf.* files were compared and combined for reference Nichols genome using MultiVCFAnalyser v. 0.87 (https://github.com/alexherbig/MultiVCFAnalyzer [177]). For each position, an allele is called if covered by a minimum of three reads, having a mapping quality of 30 and a frequency of 90% or greater among all reads covering the position. Otherwise, the position is considered as "N." All variant sites are concatenated to a SNP alignment. Additionally, a genome wide alignment with respect to the reference genome was created (Table S3).

To compare the ancient and modern treponemal genomes, a genome wide alignment and concatenated SNP alignment of *Treponema pallidum* sub. *pallidum* (n = 5), *Treponema pallidum* sub. *endemicum* (n = 2) and *Treponema pallidum* sub. *pertenue* (n = 26) genomes was generated. A maximum likelihood tree was reconstructed using RAxML [174], for 1000 replicates with a GTR model of substitution and a GAMMA distribution of rate heterogeneity model for eight categories. SNPs in SJN003 and 133 were analyzed in comparison to other treponemal genomes in order to identify private variants, variants shared between the two genomes and variants that appear to be homoplastic (Table S3). We identified regions of recombination in the genome wide alignments using ClonalFrameML ver. 1.-178 (https://github.com/xavierdidelot/ClonalFrameML [178]) given the maximum likelihood tree with default parameters. The three homoplastic SNPs observed in SJN003 and sites of recombination were excluded from the genome wide alignment and a Maximum Likelihood tree was reconstructed as described above (Figure 5).

List of HBV genomes GenBank accession numbers used in the phylogenetic analysis

AB076679, AB116084, AB453988, AY738142, GQ477499, AY934764, FJ692556, FJ692598, FJ692611, GQ161813, GQ331046, AB073858, AB033555, AB219429, AB219430, AP011089, AB073835, AB287316, AB287318, AB287320, AB287321, DQ463789, DQ463792, AB241117, DQ993686, AB111946, AB112066, AB112472, DQ089767, X75656, X75665, AB048704, AB048705, AF241411, AP011100, AP011102, AP011103, AP011106, AP011108, FJ899792, JN642140, GQ477453, GQ477455, JN642160, JN642163, JN688710, JN688711, GQ922005, HE974378, KJ470893, KJ470896, KJ470898, FJ904430, FJ904436, AB033559, AB048701, AB048702, AB188243, AB210818, AM494716, AY796031, AY902768, DQ315779, X80925, X75657, X75664, AY090458, AB116654, FJ657525, AY090455, AY311369, DQ899144, DQ899146, AB116549, X75663, AF223962, AB166850,

AB056513, AB064312, AF405706, AB059660, AB375163, AY090454, AY090457, AB486012, AY330911, AJ131571, AY781180, U46935, AJ131567, AF193863, EU155824, FJ899779, AB049609, FJ904399, AY721612, AY741797, AB270543, EU594409, AB109476, AB555496, GQ205377, HM363593, JN792922, AB562463, FJ023669, EU835241, AY934764.

List of Treponema spp. genomes used in the phylogenetic analysis

133 (from Santa Isabel, colonial Mexico [71]), A10, A12, IGU, Bosnia_A, CDC 2575, CDC-2, CDC-1, Fribourg-Blanc, Gauthier, Ghana-051, HATO, Iraq B, Kampung Dalan K363, LMNP-1, M2, M3, Nichols, Sea81-4, ERS945418, ERS945420, ERS945424, ERS945426, ERS945430, ERS945436, ERS945437, ERS945442, Samoa D, Sei Geringging K403, SS14, Mexico A, 94B (from Santa Isabel, colonial Mexico [71]).

7. Discussion

7.1 Insights gained by the work presented

The work presented in this thesis has provided insights into the population history of each of the studied regions. In Manuscript A, we extend the understanding of population history of Western Eurasia in literature by quantifying the Siberian ancestry present in different present-day and ancient populations from the region for the first time. We report the earliest evidence for this ancestry in Europe, and infer the date of introduction of Siberian ancestry in the ancient population of Bolshoy Oleni Ostrov to be roughly 4,000 yBP. Although this population was discovered at the island of Bolshoy Oleni Ostrov, it is not necessary that the introduction of Siberian ancestry in the population at 4,000 yBP took place at this geographic location. We also observe that populations with larger proportions of Siberian ancestry tend to be speakers of Uralic languages. These results have interesting implications for the spread of Uralic languages into Europe, implying that Uralic languages could have spread into Europe alongside Siberian ancestry. However, a 3,500 to 4,000 year old influx of Uralic languages into Europe would predate most linguistic estimates (Honkola et al., 2013) of the introduction of these languages to the area. Instead, linguistic analyses of the Saami language suggest that a non-Uralic linguistic substrate exists within the Saami language (Aikio, 2012), presumed to originate from earlier inhabitants of the area inhabited by the Saami. In the light of such results from linguistics, we conclude that it is unlikely that the earliest population with Siberian ancestry reported in Manuscript A was part of a population movement linked with the spread of Uralic languages into Europe, although it is possible that Uralic languages spread into Europe alongside an additional stream of Siberian ancestry. By comparing the genomes of two individuals from the Iron Age lake burial at Levänluhta to the genomes of historical and present-day Saami individuals as well as Finns, we identify a genetic shift within Finland, post-dating the Iron Age. Specifically, the Iron Age inhabitants of southern Finland appear genetically closer to the Saami than to present-day Finns. These results are in line with historical and linguistic evidence suggesting that populations related to the Saami inhabited a broader geographic range in the past, which extended up to Southern Finland.

In Manuscript B, we greatly extend the available ancient genomes from Patagonia and Tierra del Fuego, by describing and analysing genomic data from 14 ancient Fuego-Patagonian individuals. The unique geographical location of Fuego-Patagonia, at the southernmost part of South America, makes this area particularly interesting for studying human history. Being the most geographically distant location from the land connecting the North and South American continents, Fuego-Patagonia (and especially Tierra del Fuego) may have not been reached by some of the waves of people into South America until a later date, if at all. Therefore, the population history of this region could be pivotal to our understanding of population dynamics within the continent. We employed an analytical approach to determining the optimal grouping of the 14 ancient individuals, which separates each set of individuals sharing a population attribution into the minimum required number of genetically distinct groups, within the limits of our resolution. By placing the resulting groupings of ancient individuals into an admixture graph model presented in (Posth et al., 2018) we infer that ancient Fuego-Patagonian groups derive

the majority of their ancestry from a lineage most closely related to a sampled population from the Los Rieles archaeological site in Chile, and dated to 5,100 yBP. This lineage contributed to the gene pool of Fuego-Patagonian populations, genetic signatures of which remain until at least roughly 200 yBP, at the youngest temporal end of our dataset. Additionally, our results reveal a population shift happening between 5,000 and 1,100 yBP in the Kaweskar population with the introduction of an ancestry component maximised in a population that inhabited the California Channel islands 5,200 yBP. This ancestry component has been previously observed in an individual from the Cuncaicha rockshelter (4480m elevation) in Peru dated to 4,200 yBP (Posth et al., 2018), where this ancestry arrived between 9,000 and 4,200 yBP and was assumed to remain geographically restricted. Although the penetrance of this ancestry component in the sampled ancient Peruvian population was limited (2%), the inferred admixture proportion in the 1,100 yBP Kaweskar population is considerably larger (27%). There are two scenarios that could give rise to such a genetic landscape. It is possible that this ancestry spread towards Fuego-Patagonia via a more coastal route, with limited success in spreading to the Andes where it is only observed at low proportion. An alternative scenario is that this ancestry component is more pronounced in the Kaweskar than in the ancient Peruvian population due to a greater difference in the population size of the two populations. If the population size of Fuego-Patagonian hunter-gatherers was significantly lower than that of the Cuncaicha individual, then a similar number of admixture events would result in a larger proportion of ancestry replacement in the Fuego-Patagonian hunter-gatherers. Of course, these two scenarios are not mutually exclusive. We note that the inferred proportion of this ancestry component in two contemporary ancient Fuego-Patagonian populations (600 yBP) appears to follow a geographical pattern, with more northern populations having a higher proportion of this ancestry than the more southern ones. This pattern suggests that dispersal of this ancestry in the area was likely gradual, causing this ancestry to get diluted as it spread through the area, and hence unlikely to have swept through Fuego-Patagonia in a fast manner. We observe that ancient Fuego-Patagonian individuals requiring this additional minor ancestry component are also modelled as more closely related to present-day Amazonian populations than those ancient groups who lack that ancestry component. This observation has implications about the population history of present-day, hinting to present-day Amazonian populations perhaps deriving part of their ancestry from populations that also harbour a fraction of ancestry related to the 5,200 year-old California-Channel-islands populations. The demographic model we built on the rare variation of present-day reference data is consistent with those inferred from previous studies (Flegontov et al., 2019). After correction for the effective number of sites, accounting for linkage between the already sparse ancient data, the confidence intervals around the maximum likelihood placement of ancient Fuego-Patagonian individuals onto the demographic model were too broad to conclude a closer relationship to any of the present-day Native American populations, a result that could be caused by a genuine deep affinity to present-day South American populations as a result of population history events that have shaped the gene pool of South America. Indeed a comparison of the confidence intervals around the placement of Fuego-Patagonian individuals and an Argentinian mummy of similar mean genomic coverage who is dated to roughly 500 yBP reveals disproportionately large confidence intervals among the Fuego-Patagonian individuals. Finally, the observed lack of European and/or African ancestry in Fuego-Patagonian individuals that post-date the arrival of Europeans to the

continent suggest that the arrival of the effects of European contact, and perhaps also the europeans themselves, to Fuego-Patagonia may have been delayed compared to the rest of the continent.

In Manuscript C, a multitude of approaches are combined to reconstruct the life story of three presumed slaves from a mass grave in Mexico City. Dental modifications suggested these individuals to be of African origin, a hypothesis which we confirmed with archaeogenetic analysis of immunogenetic markers as well as mitochondrial-DNA and Y-chromosome haplotypes. Archaeogenetic analysis of the human nuclear aDNA extracted from these individuals suggests that each individual is most closely related to a different West African population from which they appear to exclusively derive ancestry, and that none of the three individuals have received any ancestry from non-African sources. These results indicate that these individuals are genetically African, but the genetic results alone cannot prove these individuals are first-generation Africans in the New world. They could, instead, belong to a later generation of Africans born in the New world to genetically African parents. We compared the genetic affinity of each individual to other African populations with that of their respective most closely related population. Results from this analysis revealed that none of these individuals was more closely related to any tested African population than their suggested source population. Therefore, if these individuals did belong to a later generation of Africans in the New World, their parents would need to be from the same or from closely related populations. While we cannot rule out such a possibility from the genetic data alone, it remains an unlikely scenario because forced social intermixing was often used as a means of social control and disintegration among the enslaved population (Barquera & Acuña-Alonzo, 2012; Hurston, 2018). Strontium ratios ($^{87}Sr/^{86}Sr$) obtained from the molars of these individuals, which form between the 2nd and 8th years of their lives, gives ratios inconsistent with any ratios recorded in Mexico, implying that these individuals did not spend their childhood in Mexico. Instead, these ratios are consistent with ratios observed in Western Africa. In concert, these results confirm these individuals as first-generation migrants to Mexico. The phylogenetic analysis of the recovered *Treponema pallidum sub. pertenue* (the causative agent of yaws) and Hepatitis B Virus (HBV) genomes revealed a genetic similarity with strains of either pathogenic agent presently found in Western Africa. Together, these results shed light into the effects the transatlantic slave-trade has had in the dissemination of disease across the world. The osteological analysis of the bones identified signs of poor hygiene, infectious processes, and malnutrition, as well as signatures of conflict and hardship, results consistent with the hypothesis that these individuals were forcibly migrated to Mexico as part of the slave-trade. The extent of these hardships is visible on the remains. As examples of the physical hardships of these individuals, the remains of one individual (SJN001) show signs of this individual having been shot (López Wario et al., 1996) and buried with the fragments still in their body, while the remains of another individual (SJN002) show signs of broken bones in their lower leg (Espinoza et al., 2015) and a cut mark on the cranium. By studying the life history and death of individuals that suffered during times of suppression and hardship, we attempt to immortalise these parts of their experience. The hardships that these individuals were subjected to left marks that are still visible today, on both their own remains and the world at large.

7.2 Limitations to the archaeogenetic approach

Archaeogenetics has proven to be a formidable approach to studying human history. Sampling ancient populations across temporal and spatial gradients can provide many insights into the movement of new populations into an area, and admixture between populations. However, this approach is not without its limitations.

Differences in preservation and discovery of ancient biological material make it impossible to ever achieve perfect temporal and spatial sampling of ancient populations. Additionally, incomplete preservation of aDNA due to environmental factors introduces varying degrees of missing data to archaeogenetic analyses and often limits the genomic coverage that can be achieved at reasonable monetary expense. Technological advancements in retrieval (Meyer et al., 2012; Pinhasi et al., 2015), and enrichment (Fu, Meyer, et al., 2013) of aDNA from archaeological material have drastically improved success rates for sampling of ancient material, while advances in sequencing technologies have decreased the costs of sequencing (Margulies et al., 2005). Although technological and technical innovations continuously improve our ability to sample ancient populations, bioarchaeological material is a limited resource and complete spatiotemporal sampling of ancient populations is not a realistic possibility currently.

Finally, some information about the studied individuals and populations cannot be retrieved using archaeogenetics, such as cultural identity, spoken language, and subsistence strategy. To overcome these limitations it is important to combine archaeogenetics with other approaches to the study of human history. For example, in Manuscript C, although archaeogenetic analyses prove that all the studied individuals are of West African ancestry, this approach cannot distinguish if these individuals were born in Africa or if they were born in the Americas. In contrast, with the recovered strontium ratios it is possible to disprove a local origin for these individuals, but not to identify their origin. Only by combining the two approaches were we able to identify these individuals as first-generation Africans.

7.3 How many individuals is enough?

By providing direct genetic evidence for gene flow between different (sub-)species of humans (Green et al., 2010; Meyer et al., 2012), migration of people, and admixture between different human populations (Allentoft et al., 2015; Haak et al., 2015; Lazaridis et al., 2014; Mathieson et al., 2018; Raghavan et al., 2014), the field of archaeogenetics has significantly altered the way people think about population history by highlighting the role that the movement of genes has had in it. According to our current understanding, human population history is less like a tree and more like a web, with frequent admixture events connecting different lineages together. From this perspective, any population, present-day or ancient, can be described as a mixture of mixtures of different ancestries, with the addition of varying degrees of drift accumulating on each lineage over time. This abstracted view of a population as a collection of different distal ancestries is a useful model for genetic studies, and one that comes in stark contrast with many of the narratives put forward through the study of human populations in the previous centuries. In those narratives, a person's ancestry is often defined in terms of a single source, while ignoring the intricate genetic relationships we now know these sources have with

one another. Misconceptions borne of these flawed narratives can still be encountered today, especially with the advent of personal ancestry testing services, when users of such services often equate having a proportion of their genome be most closely related to present-day populations from an area with deriving ancestry from an ancient population from that area (e.g. sharing ancestry with Scandinavian populations does not necessarily mean you descend from a Viking). As more ancient populations are sampled to fill the remaining spatiotemporal gaps, the resolution of comparative studies will increase, eventually explaining as mixtures those lineages that cannot be explained as a mixture of distal sources with the currently available data.

By making use of technological advancements in the field, the number of newly published ancient genomes per archaeogenetic study has increased dramatically in recent years. Figure C shows the maximum number of newly published aDNA genomes in a single study between 2012 and 2019. There are numerous scientific reasons for this trend. Broader geographical and temporal sampling adds to the resolution of comparative analyses, as mentioned above. Additionally, larger sample sizes per population can overcome some of the limitations of current archaeogenetic methods, like F- and D-statistics. As already explained, these statistics utilise the allele frequency differences between different populations at a set of SNPs. These allele frequency estimates are a direct consequence of the specific genotype calls all individuals of a population have. A common approach to genotyping ancient individuals on a set of desired SNPs is to randomly sample a single sequencing read covering the position of each SNP, and make a homozygous call with the allele present on this read for that SNP and individual. While this approach might seem counter-intuitive at first glance, because it only utilises a fraction of the total sequencing data that has been generated, it is effectively an assumption-free approach to genotyping, thus reducing reference bias, which could otherwise affect population genetic results (Günther & Nettelblad, 2019). The downside of this approach is that every genotyped individual is effectively haploid, thus halving the number of chromosomes used when estimating allele frequencies per SNP and resulting in more coarse estimates. The coarser allele frequency estimate per SNP will affect the power of F- and D-statistics through the addition of noise, without introducing systematic directional changes in their values. Furthermore, the resolution of these analytical methods differ by the specific context they are applied in. Allele frequency methods, like the F- and D-statistics can confidently identify even low levels of admixture in cases where the allele frequencies of the source populations are very different (e.g. Neanderthal introgression) (Green et al., 2010), but have limited resolution in cases where the two sources have similar allele frequencies, especially if the contribution of one of these sources is small. There are two ways to overcome this limitation: a) sample enough individuals in each population to reduce the errors around allele frequency estimates, thus increasing resolution, or b) use larger/different portions of the site frequency spectrum, to leverage more of the genetic variation between human populations.

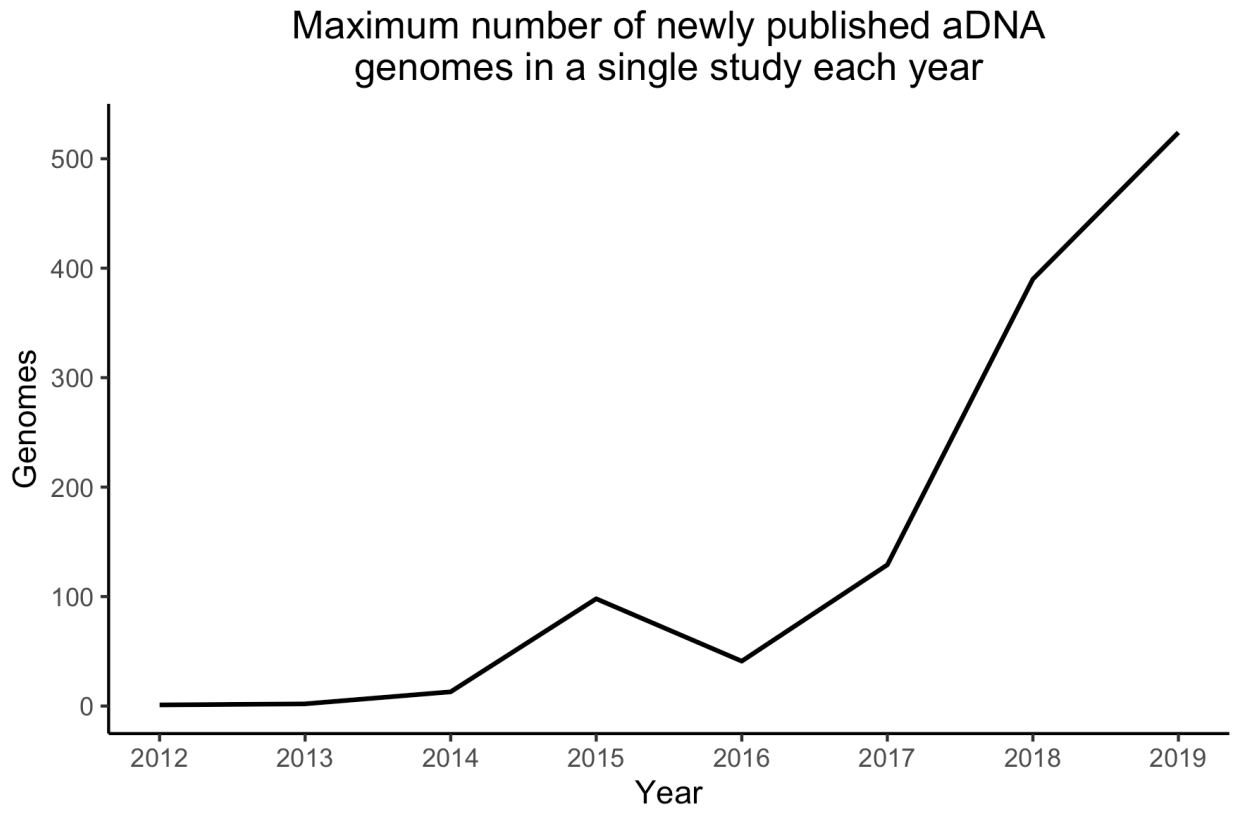


Figure C. The maximum number of newly published ancient genomes in a single study per year from 2012-2019. Data from: (*Downloadable genotypes of present-day and ancient DNA data (compiled from published papers)* | *David Reich Lab*, n.d.) (v42.4)

Although more extensive sampling can overcome these issues, one must consider the cost that extensive sampling comes with. Archaeological material is limited, and sampling human bones for archaeogenetic studies is a destructive process. While the amount of bone powder required for DNA extraction is small (30-50mg) (I. Velsko et al., n.d.) compared to other destructive analyses, such as radiocarbon dating which often requires upwards of 150mg of material, the material preferentially used for aDNA analysis (i.e. teeth and/or petrous bones) is a small subset of all available bone material, and therefore even more limited. Additionally, although technological advancements like in-solution capture have increased success rates for aDNA extraction and sequencing, these technologies are not yet widely available. As sampling, extraction, and sequencing techniques continue to develop, success rates generation of high-quality aDNA data from human remains will continue to increase in the future. This poses an interesting dilemma: Should we risk destroying precious archaeological material now in hopes of successfully analysing it, or wait for methods to evolve until the associated risks are lowered first? To some extent, these risks can be lessened by the safekeeping of bone powder and immortalised DNA libraries created from the extracted ancient DNA when possible, which can be used to re-extract DNA and/or sequence the genome of an ancient individual further in the future, but the risks can never be completely overcome. For example, advances in sampling technologies cannot be applied to pre-existing DNA libraries.

The alternative solution of utilising more of the site frequency spectrum requires the use, adaptation, and development of more analytical methods. One step in this direction is to use and/or adapt methods that are already used in population genetics of present-day populations. Conceptually, this is a more proximal solution but not one without challenges. Whereas genetic studies of present-day populations will readily exclude individuals with high missingness rates from their analyses, ancient genomes will regularly be included in analyses even when missing genotypes for 50% of the SNPs in the comparative dataset. In fact, missingness rates close to 90% are tolerated for some analyses, although the results are rarely conclusive in extreme cases. Before existing methods can be used with ancient data, it is important to first test their robustness to missing data, or adapt them to deal with this challenge (Lamnidis et al., 2018; Skoglund et al., 2012). Imputation of ancient haplotypes based on a reference dataset of present-day haplotypes is also a promising approach (Martiniano et al., 2017), which can help reduce missingness while also resulting in much higher resolution data. Another promising approach is the analysis of shared rare variation between ancient and present-day populations. While this approach is currently only economically viable for individuals with higher DNA preservation, and often requires a large sequencing effort per ancient individual, this approach has been successfully used to study British migration history during the Iron Age (Schiffels et al., 2016) as well as the populations of North America and Northeastern Siberia (Flegontov et al., 2019). The former is a time where allele frequency differences between European populations are smaller since most West Eruasian populations at the time derive their ancestry from the same three ancestral populations, thus limiting the resolution of analytical methods dependent on the magnitude of allele frequency differences between populations.

Another reason for this upward trend in the number of individuals analysed in a single archaeogenetic study has to do with the questions each study is attempting to answer. A limited number of ancient individuals carrying an ancestry component that was thus far unsampled can provide many valuable insights into the population history of a broad spatiotemporal area (Raghavan et al., 2014). This is especially true when these ancient individuals are analysed against a reference set of present-day populations who partly carry ancestries related to that of the analysed ancient individuals, as is the case in Manuscript A (Lamnidis et al., 2018). But as the spatiotemporal gaps in available ancient genomes become smaller, the large-scale insights a handful of genomes can reveal become more limited. Instead, many archaeogenetic studies nowadays focus on discerning sociocultural aspects of the studied populations, such as male/female exogamy, kinship patterns across different burials within an archaeological site, and/or the extent of inbreeding in the population. For instance, observing more males being buried in the same area as their genetic relatives than females on average, we can infer that the population(s) using these burial grounds were likely practicing patrilocality (Furtwängler et al., 2020; Mittnik et al., 2019; Sánchez-Quinto et al., 2019). Recently, signs of genetic inbreeding were discovered in the genome of an individual from an elaborate recess tomb in the Neolithic archaeological tomb of Newgrange in Ireland, which suggests the existence of social stratification and political integration in the Neolithic societies of Ireland (Cassidy et al., 2020), because in most human societies inbreeding is considered taboo except among members of a social elite (e.g. royal families). Because such inferences of sociocultural traits of a population rely on identifying genetic connections between ancient individuals, they benefit greatly from increased sampling coverage of an archaeological site, which allows more of these connections

to be identified, while also painting a more complete picture of the archaeological site. As the spatiotemporal sampling coverage of a region increases, we can expect to also detect kinship relationships between archaeological sites. For example, identifying relatives of a non-local individual in another archaeological site would shed much light into the interactions of the populations using these burial sites, and will perhaps allow future researchers to build networks of population movement between different settlements in a geographical area based on the genetic kinship of individuals across different burial sites.

7.4 Thoughts about the future of thinking about the past

Thus far, the majority of archaeogenetic studies have looked at continental patterns of genetic variation, admixture and migration. But as discussed above, more complete sampling of an archaeological site to identify genetic kinships, and therefore social structure and mating patterns of a population is becoming an increasingly popular approach to archaeogenetic analysis. I believe the future of archaeogenetics will include a lot more interdisciplinary analysis like the work presented in Manuscript C, to reconstruct many facets of the life and death of the analysed ancient individuals. Through such studies, a much deeper understanding of the sociocultural dynamics of the analysed individuals will be possible, and they will (re)gain an identity beyond the museum or laboratory IDs that were assigned to their remains. Learning a lot of information about three specific individuals has made their story much more personal to me, both as a researcher having worked towards uncovering facets of this story, but also as a reader of their story. Being confronted with the story of hardship, violence and poor living conditions of people who survived the trans-atlantic slave trade (as told from their own remains) is emotionally very different to simply reading statistics or general statements regarding these atrocities. It is intriguing to ponder how this type of knowledge might change how we think about the present and the past. There are two, in my view positive, changes that can be expected.

The first change concerns scientific outreach. Personal narratives like the ones in Manuscript C are often used in museum exhibitions in an attempt to educate the public and get them more involved with research of human history (e.g. the shaman woman of Bad Dürrenberg). As aDNA research becomes more ubiquitous, and the knowledge required for extraction, sequencing and analysis thereof becomes more accessible, I expect we will see a rise in the number of personal narratives including results from archaeogenetic analysis of the material in question. Engagement with such literature can have various social knock-on benefits that are outside the scope of this thesis, but I would like to briefly mention one. Through engaging with archaeogenetics literature, and learning about the view of populations as mixtures of mixtures, the public can begin to better understand how concepts like “race” are cultural constructs, and perhaps eventually dispel the connection between cultural and genetic identity, a topic that recurrently comes up in public discourse.

The second change concerns archaeogeneticists ourselves. While the study of human population history will always be centered around population-level events, such as migration and admixture, it is important to also remember that human populations are made up of individuals. Each individual has their own personal history, exhibited certain cultural traits, was part of a society (often with its own social structure), died of different causes, and had specific

dietary habits. As such, it is important for us to strive to tell these individual stories as well when possible, and to try to get the most information possible out of ancient individuals if we choose to sample their remains. This might entail not only analysing the extracted human aDNA, but also screening for aDNA of pathogenic agents, additionally sampling the dental calculus of an individual to reconstruct their oral microbiome, carrying out dietary isotope analysis, radiocarbon dating, and more. As shown already with Manuscript C, there is a lot we stand to learn when we combine all the tools in our toolkit.

8. References

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Aikio, A. (2012). An essay on Saami ethnolinguistic prehistory. *Suomalais-Ugrilaisen Seuran Toimituksia = Mémoires de La Société Finno-Ougrienne*, *266*, 63–117. http://www.sgr.fi/sust/sust266/sust266_aikio.pdf

Allentoft, M. E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P. B., Schroeder, H., Ahlström, T., Vinner, L., Malaspinas, A.-S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Della Casa, P., Dąbrowski, P., … Willerslev, E. (2015). Population genomics of Bronze Age Eurasia. *Nature*, *522*(7555), 167–172. https://doi.org/10.1038/nature14507

Anagnostou, P., Capocasa, M., Milia, N., Sanna, E., Battaggia, C., Luzi, D., & Destro Bisol, G. (2015). When data sharing gets close to 100%: what human paleogenetics can teach the open science movement. *PloS One*, *10*(3), e0121409. https://doi.org/10.1371/journal.pone.0121409

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., & Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, *290*(5806), 457–465. https://doi.org/10.1038/290457a0

Andrades Valtueña, A., Mittnik, A., Key, F. M., Haak, W., Allmäe, R., Belinskij, A., Daubaras, M., Feldman, M., Jankauskas, R., Janković, I., Massy, K., Novak, M., Pfrengle, S.,

Reinhold, S., Šlaus, M., Spyrou, M. A., Szécsényi-Nagy, A., Tõrv, M., Hansen, S., … Krause, J. (2017). The Stone Age Plague and Its Persistence in Eurasia. *Current Biology: CB*, *27*(23), 3683–3691.e8. https://doi.org/10.1016/j.cub.2017.10.025

Barker, G. (2006). *The Agricultural Revolution in Prehistory: Why did Foragers become Farmers?* OUP Oxford. https://play.google.com/store/books/details?id=fkifXu2gx4YC

Barquera, R., & Acuña-Alonzo, V. (2012). The African colonial migration into Mexico: history and biological consequences. In *Causes and Consequences of Human Migration: An Evolutionary Perspective* (pp. 201–223). Cambridge University Press. https://doi.org/10.1017/CBO9781139003308.014

Bos, K. I., Herbig, A., Sahl, J., Waglechner, N., Fourment, M., Forrest, S. A., Klunk, J., Schuenemann, V. J., Poinar, D., Kuch, M., Golding, G. B., Dutour, O., Keim, P., Wagner, D. M., Holmes, E. C., Krause, J., & Poinar, H. N. (2016). Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. *eLife*, *5*, e12994. https://doi.org/10.7554/eLife.12994

Bunn, H. F. (2013). The triumph of good over evil: protection by the sickle gene against malaria. *Blood*, *121*(1), 20–25. https://doi.org/10.1182/blood-2012-08-449397

Cassidy, L. M., Maoldúin, R. Ó., Kador, T., Lynch, A., Jones, C., Woodman, P. C., Murphy, E., Ramsey, G., Dowd, M., Noonan, A., Campbell, C., Jones, E. R., Mattiangeli, V., & Bradley, D. G. (2020). A dynastic elite in monumental Neolithic society. *Nature*, *582*(7812), 384–388. https://doi.org/10.1038/s41586-020-2378-6

Der Sarkissian, C., Ermini, L., Schubert, M., Yang, M. A., Librado, P., Fumagalli, M., Jónsson, H., Bar-Gal, G. K., Albrechtsen, A., Vieira, F. G., Petersen, B., Ginolhac, A., Seguin-Orlando, A., Magnussen, K., Fages, A., Gamba, C., Lorente-Galdos, B., Polani, S., Steiner,

C., … Orlando, L. (2015). Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Current Biology: CB*, *25*(19), 2577–2583. https://doi.org/10.1016/j.cub.2015.08.032

Downloadable genotypes of present-day and ancient DNA data (compiled from published papers) | David Reich Lab. (n.d.). Retrieved June 26, 2020, from https://reich.hms.harvard.edu/downloadable-genotypes-present-day-and-ancient-dna-data-compiled-published-papers

Dunbar-Ortiz, R. (2014). *An Indigenous Peoples' History of the United States*. Beacon Press. https://play.google.com/store/books/details?id=ZkEoAwAAQBAJ

Espinoza, P. O. H., Lagunas, Z. R., & Others. (2015). *Manual de Osteología*. Instituto Nacional de Antropología e Historia/Escuela Nacional de ….

Flegontov, P., Altınışık, N. E., Changmai, P., Rohland, N., Mallick, S., Adamski, N., Bolnick, D. A., Broomandkhoshbacht, N., Candilio, F., Culleton, B. J., Flegontova, O., Friesen, T. M., Jeong, C., Harper, T. K., Keating, D., Kennett, D. J., Kim, A. M., Lamnidis, T. C., Lawson, A. M., … Schiffels, S. (2019). Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature*, *570*(7760), 236–240. https://doi.org/10.1038/s41586-019-1251-y

Friedlaender, J., Schurr, T., Gentz, F., Koki, G., Friedlaender, F., Horvat, G., Babb, P., Cerchio, S., Kaestle, F., Schanfield, M., Deka, R., Yanagihara, R., & Merriwether, D. A. (2005). Expanding Southwest Pacific mitochondrial haplogroups P and Q. *Molecular Biology and Evolution*, *22*(6), 1506–1517. https://doi.org/10.1093/molbev/msi142

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H. A., Kelso, J., & Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the*

National Academy of Sciences of the United States of America, *110*(6), 2223–2227. https://doi.org/10.1073/pnas.1221359110

Fu, Q., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., Ronchitelli, A. M., Martini, F., Cremonesi, R. G., Svoboda, J., Bauer, P., Caramelli, D., Castellano, S., Reich, D., Pääbo, S., & Krause, J. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology: CB*, *23*(7), 553–559. https://doi.org/10.1016/j.cub.2013.02.044

Furtwängler, A., Rohrlach, A. B., Lamnidis, T. C., Papac, L., Neumann, G. U., Siebke, I., Reiter, E., Steuri, N., Hald, J., Denaire, A., Schnitzler, B., Wahl, J., Ramstein, M., Schuenemann, V. J., Stockhammer, P. W., Hafner, A., Lösch, S., Haak, W., Schiffels, S., & Krause, J. (2020). Ancient genomes reveal social and genetic structure of Late Neolithic Switzerland. *Nature Communications*, *11*(1), 1915. https://doi.org/10.1038/s41467-020-15560-x

Graur, D. (2017). An Upper Limit on the Functional Fraction of the Human Genome. *Genome Biology and Evolution*, *9*(7), 1880–1885. https://doi.org/10.1093/gbe/evx121

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Green, R. E., Malaspinas, A.-S., Krause, J., Briggs, A. W., Johnson, P. L. F., Uhler, C., Meyer, M., Good, J. M., Maricic, T., Stenzel, U., Prüfer, K., Siebauer, M., Burbano, H. A., Ronan, M., Rothberg, J. M., Egholm, M., Rudan, P., Brajković, D., Kućan, Z., … Pääbo, S. (2008). A complete Neandertal mitochondrial genome sequence determined by high-throughput

sequencing. *Cell*, *134*(3), 416–426. https://doi.org/10.1016/j.cell.2008.06.021

Gretzinger, J., Molak, M., Reiter, E., Pfrengle, S., Urban, C., Neukamm, J., Blant, M., Conard, N. J., Cupillard, C., Dimitrijević, V., Drucker, D. G., Hofman-Kamińska, E., Kowalczyk, R., Krajcarz, M. T., Krajcarz, M., Münzel, S. C., Peresani, M., Romandini, M., Rufí, I., … Schuenemann, V. J. (2019). Large-scale mitogenomic analysis of the phylogeography of the Late Pleistocene cave bear. *Scientific Reports*, *9*(1), 10700. https://doi.org/10.1038/s41598-019-47073-z

Günther, T., & Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genetics*, *15*(7), e1008302. https://doi.org/10.1371/journal.pgen.1008302

Haak, W., Balanovsky, O., Sanchez, J. J., Koshel, S., Zaporozhchenko, V., Adler, C. J., Der Sarkissian, C. S. I., Brandt, G., Schwarz, C., Nicklisch, N., Dresely, V., Fritsch, B., Balanovska, E., Villems, R., Meller, H., Alt, K. W., Cooper, A., & Members of the Genographic Consortium. (2010). Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biology*, *8*(11), e1000536. https://doi.org/10.1371/journal.pbio.1000536

Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D., Alt, K. W., & Burger, J. (2005). Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science*, *310*(5750), 1016–1018. https://doi.org/10.1126/science.1118725

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bánffy, E., Economou, C., Francken, M., Friederich, S., Pena, R. G., Hallgren, F., Khartanovich, V., … Reich, D.

(2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, *522*(7555), 207–211. https://doi.org/10.1038/nature14317

Handt, O., Höss, M., Krings, M., & Pääbo, S. (1994). Ancient DNA: methodological challenges. *Experientia*, *50*(6), 524–529. https://doi.org/10.1007/bf01921720

Hedges, S. B., & Schweitzer, M. H. (1995). Detecting dinosaur DNA [Review of *Detecting dinosaur DNA*]. *Science*, *268*(5214), 1191–1192; author reply 1194. https://doi.org/10.1126/science.7761839

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A., & Wilson, A. C. (1984). DNA sequences from the quagga, an extinct member of the horse family. *Nature*, *312*(5991), 282–284. https://doi.org/10.1038/312282a0

Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-Del-Molino, D., van Dorp, L., López, S., Kousathanas, A., Link, V., Kirsanow, K., Cassidy, L. M., Martiniano, R., Strobel, M., Scheu, A., Kotsakis, K., Halstead, P., Triantaphyllou, S., Kyparissi-Apostolika, N., … Burger, J. (2016). Early farmers from across Europe directly descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(25), 6886–6891. https://doi.org/10.1073/pnas.1523951113

Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, A., & Pääbo, S. (2001). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*, *29*(23), 4793–4799. https://www.ncbi.nlm.nih.gov/pubmed/11726688

Honkola, T., Vesakoski, O., Korhonen, K., Lehtinen, J., Syrjänen, K., & Wahlberg, N. (2013). Cultural and climatic changes shape the evolutionary history of the Uralic languages. *Journal of Evolutionary Biology*, *26*(6), 1244–1253. https://doi.org/10.1111/jeb.12107

Hurston, Z. N. (2018). *Barracoon: The Story of the Last" black Cargo."* HarperCollins.

Ikehata, H., & Ono, T. (2011). The mechanisms of UV mutagenesis. *Journal of Radiation Research*, *52*(2), 115–125. https://doi.org/10.1269/jrr.10175

Ilumäe, A.-M., Reidla, M., Chukhryaeva, M., Järve, M., Post, H., Karmin, M., Saag, L., Agdzhoyan, A., Kushniarevich, A., Litvinov, S., Ekomasova, N., Tambets, K., Metspalu, E., Khusainova, R., Yunusbayev, B., Khusnutdinova, E. K., Osipova, L. P., Fedorova, S., Utevska, O., … Rootsi, S. (2016). Human Y Chromosome Haplogroup N: A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families. *American Journal of Human Genetics*, *99*(1), 163–173. https://doi.org/10.1016/j.ajhg.2016.05.025

Isern, N., Fort, J., & Vander Linden, M. (2012). Space competition and time delays in human range expansions. Application to the neolithic transition. *PloS One*, *7*(12), e51106. https://doi.org/10.1371/journal.pone.0051106

Jones, E. R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R. L., Gallego Llorente, M., Cassidy, L. M., Gamba, C., Meshveliani, T., Bar-Yosef, O., Müller, W., Belfer-Cohen, A., Matskevich, Z., Jakeli, N., Higham, T. F. G., Currat, M., Lordkipanidze, D., … Bradley, D. G. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature Communications*, *6*, 8912. https://doi.org/10.1038/ncomms9912

Kistler, L., Ware, R., Smith, O., Collins, M., & Allaby, R. G. (2017). A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Research*, *45*(11), 6310–6320. https://doi.org/10.1093/nar/gkx361

Kivisild, T. (2015). Maternal ancestry and population history from whole mitochondrial genomes. *Investigative Genetics*, *6*, 3. https://doi.org/10.1186/s13323-015-0022-2

Kılınç, G. M., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, A. M., Bıçakçı, E., Baird, D., Dönertaş, H. M., Ghalichi, A., Yaka, R., Koptekin, D., Açan, S. C., Parvizi, P., Krzewińska, M., Daskalaki, E. A., Yüncü, E., Dağtaş, N. D., Fairbairn, A., Pearson, J., … Götherström, A. (2016). The Demographic Development of the First Farmers in Anatolia. *Current Biology: CB*, *26*(19), 2659–2666. https://doi.org/10.1016/j.cub.2016.07.057

Lamnidis, T. C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., Khartanovich, V., Balanovsky, O., Ongyerth, M., Weihmann, A., Sajantila, A., Kelso, J., Pääbo, S., Onkamo, P., Haak, W., Krause, J., & Schiffels, S. (2018). Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nature Communications*, *9*(1), 5018. https://doi.org/10.1038/s41467-018-07483-5

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D. C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., Connell, S., Stewardson, K., Harney, E., Fu, Q., Gonzalez-Fortes, G., Jones, E. R., Roodenberg, S. A., Lengyel, G., Bocquentin, F., … Reich, D. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature*, *536*(7617), 419–424. https://doi.org/10.1038/nature19310

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P. H., Schraiber, J. G., Castellano, S., Lipson, M., Berger, B., Economou, C., Bollongino, R., Fu, Q., Bos, K. I., Nordenfelt, S., Li, H., de Filippo, C., Prüfer, K., … Krause, J. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, *513*(7518), 409–413. https://doi.org/10.1038/nature13673

Lindo, J., Haas, R., Hofman, C., Apata, M., Moraga, M., Verdugo, R. A., Watson, J. T., Viviano Llave, C., Witonsky, D., Beall, C., Warinner, C., Novembre, J., Aldenderfer, M., & Di Rienzo, A. (2018). The genetic prehistory of the Andean highlands 7000 years BP though

European contact. *Science Advances*, *4*(11), eaau4921. https://doi.org/10.1126/sciadv.aau4921

Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., Oppenheimer, J., Broomandkhoshbacht, N., Harney, E., Nordenfelt, S., Llamas, B., Gusztáv Mende, B., Köhler, K., Oross, K., Bondár, M., … Reich, D. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature*, *551*(7680), 368–372. https://doi.org/10.1038/nature24476

López Wario, L. A., Meza Peñaloza, A., & Báez-Molgado, S. (1996). Una muerte violenta en el virreinato (el caso del esqueleto 150 de la línea 8 del Metro, México, D.F.). *Rev. La Coord. Nac. Arqueol. Del Inst. Nac. Antropol. E Hist. Ene-Jun*, 111–114. https://www.scopus.com/record/display.uri?eid=2-s2.0-85085750185&origin=inward

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., Skoglund, P., Lazaridis, I., Sankararaman, S., Fu, Q., Rohland, N., Renaud, G., Erlich, Y., Willems, T., Gallo, C., … Reich, D. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, *538*(7624), 201–206. https://doi.org/10.1038/nature18964

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., … Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057), 376–380. https://doi.org/10.1038/nature03959

Martiniano, R., Cassidy, L. M., Ó'Maoldúin, R., McLaughlin, R., Silva, N. M., Manco, L., Fidalgo, D., Pereira, T., Coelho, M. J., Serra, M., Burger, J., Parreira, R., Moran, E., Valera,

A. C., Porfirio, E., Boaventura, R., Silva, A. M., & Bradley, D. G. (2017). The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genetics*, *13*(7), e1006852. https://doi.org/10.1371/journal.pgen.1006852

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., Fernandes, D., Ferry, M., Gamarra, B., Fortes, G. G., Haak, W., Harney, E., Jones, E., Keating, D., Krause-Kyora, B., … Reich, D. (2018). The genomic history of southeastern Europe. *Nature*, *555*(7695), 197–203. https://doi.org/10.1038/nature25778

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E. R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J. L., de Castro, J. M. B., Carbonell, E., … Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*(7583), 499–503. https://doi.org/10.1038/nature16152

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., de Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R. E., Bryc, K., … Pääbo, S. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science*, *338*(6104), 222–226. https://doi.org/10.1126/science.1224344

Mittnik, A., Massy, K., Knipper, C., Wittenborn, F., Friedrich, R., Pfrengle, S., Burri, M., Carlichi-Witjes, N., Deeg, H., Furtwängler, A., Harbeck, M., von Heyking, K., Kociumaka, C., Kucukkalipci, I., Lindauer, S., Metz, S., Staskiewicz, A., Thiel, A., Wahl, J., … Krause, J. (2019). Kinship-based social inequality in Bronze Age Europe. *Science*, *366*(6466), 731–

734. https://doi.org/10.1126/science.aax6219

Montinaro, F., Busby, G. B. J., Pascali, V. L., Myers, S., Hellenthal, G., & Capelli, C. (2015). Unravelling the hidden ancestry of American admixed populations. *Nature Communications*, *6*, 6596. https://doi.org/10.1038/ncomms7596

Moreno-Mayar, J. V., Potter, B. A., Vinner, L., Steinrücken, M., Rasmussen, S., Terhorst, J., Kamm, J. A., Albrechtsen, A., Malaspinas, A.-S., Sikora, M., Reuther, J. D., Irish, J. D., Malhi, R. S., Orlando, L., Song, Y. S., Nielsen, R., Meltzer, D. J., & Willerslev, E. (2018). Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature*. https://doi.org/10.1038/nature25173

Moreno-Mayar, J. V., Vinner, L., de Barros Damgaard, P., de la Fuente, C., Chan, J., Spence, J. P., Allentoft, M. E., Vimala, T., Racimo, F., Pinotti, T., Rasmussen, S., Margaryan, A., Iraeta Orbegozo, M., Mylopotamitaki, D., Wooller, M., Bataille, C., Becerra-Valdivia, L., Chivall, D., Comeskey, D., … Willerslev, E. (2018). Early human dispersals within the Americas. *Science*. https://doi.org/10.1126/science.aav2621

Morgunova, N., & Khokhlova, O. (2013). Chronology and Periodization of the Pit-Grave Culture in the Area Between the Volga and Ural Rivers Based on 14C Dating and Paleopedological Research. *Radiocarbon*, *55*(2–3), 1286–1296. https://journals.uair.arizona.edu/index.php/radiocarbon/article/view/16087

Mullis, K. B., & Faloona, F. A. (1989). 9 - Specific Synthesis of DNA in Vitro via a Polymerase-Catalyzed Chain Reaction. In R. Wu, L. Grossman, & K. Moldave (Eds.), *Recombinant DNA Methodology* (pp. 189–204). Academic Press. https://doi.org/10.1016/B978-0-12-765560-4.50015-0

Narasimhan, V. M., Rahbari, R., Scally, A., Wuster, A., Mason, D., Xue, Y., Wright, J.,

Trembath, R. C., Maher, E. R., van Heel, D. A., Auton, A., Hurles, M. E., Tyler-Smith, C., & Durbin, R. (2017). Estimating the human mutation rate from autozygous segments reveals population differences in human mutational processes. *Nature Communications*, *8*(1), 303. https://doi.org/10.1038/s41467-017-00323-y

Nesheva, D. (2014). Aspects of ancient mitochondrial DNA analysis in different populations for understanding human evolution. *Balkan Journal of Medical Genetics: BJMG*, *17*(1), 5–14. https://doi.org/10.2478/bjmg-2014-0019

O'Fallon, B. D., & Fehren-Schmitz, L. (2011). Native Americans experienced a strong population bottleneck coincident with European contact. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(51), 20444–20448. https://doi.org/10.1073/pnas.1112563108

Olalde, I., Schroeder, H., Sandoval-Velasco, M., Vinner, L., Lobón, I., Ramirez, O., Civit, S., García Borja, P., Salazar-García, D. C., Talamo, S., María Fullola, J., Xavier Oms, F., Pedro, M., Martínez, P., Sanz, M., Daura, J., Zilhão, J., Marquès-Bonet, T., Gilbert, M. T. P., & Lalueza-Fox, C. (2015). A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. *Molecular Biology and Evolution*, *32*(12), 3132–3142. https://doi.org/10.1093/molbev/msv181

Orlando, L., Bonjean, D., Bocherens, H., Thenot, A., Argant, A., Otte, M., & Hänni, C. (2002). Ancient DNA and the population genetics of cave bears (Ursus spelaeus) through space and time. *Molecular Biology and Evolution*, *19*(11), 1920–1933. https://doi.org/10.1093/oxfordjournals.molbev.a004016

Orlando, L., Male, D., Alberdi, M. T., Prado, J. L., Prieto, A., Cooper, A., & Hänni, C. (2008). Ancient DNA clarifies the evolutionary history of American Late Pleistocene equids.

Journal of Molecular Evolution, *66*(5), 533–538. https://doi.org/10.1007/s00239-008-9100-x

Pääbo, S. (1984). Über den Nachweis von DNA in altägyptischen Mumien. *Das Altertum*, *30*, 213–218.

Pääbo, S., Higuchi, R. G., & Wilson, A. C. (1989). Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *The Journal of Biological Chemistry*, *264*(17), 9709–9712. https://www.ncbi.nlm.nih.gov/pubmed/2656708

Pääbo, S., & Wilson, A. C. (1988). Polymerase chain reaction reveals cloning artefacts. *Nature*, *334*(6181), 387–388. https://doi.org/10.1038/334387b0

Palkopoulou, E., Lipson, M., Mallick, S., Nielsen, S., Rohland, N., Baleka, S., Karpinski, E., Ivancevic, A. M., To, T.-H., Kortschak, R. D., Raison, J. M., Qu, Z., Chin, T.-J., Alt, K. W., Claesson, S., Dalén, L., MacPhee, R. D. E., Meller, H., Roca, A. L., … Reich, D. (2018). A comprehensive genomic history of extinct and living elephants. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), E2566–E2574. https://doi.org/10.1073/pnas.1720554115

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., Mendoza, M. L. Z., Beaudoin, A. B., Zutter, C., Larsen, N. K., Potter, B. A., Nielsen, R., Rainville, R. A., Orlando, L., Meltzer, D. J., Kjær, K. H., & Willerslev, E. (2016). Postglacial viability and colonization in North America's ice-free corridor. *Nature*, *537*(7618), 45–49. https://doi.org/10.1038/nature19085

Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczky, P., Anders, A., Pietrusewsky, M., Rollefson, G., Jovanovic, M., Trinhhoang, H., Bar-Oz, G., Oxenham, M., Matsumura, H., & Hofreiter, M. (2015). Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PloS One*, *10*(6), e0129102. https://doi.org/10.1371/journal.pone.0129102

Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., Lamnidis, T. C., Rohland, N., Nägele, K., Adamski, N., Bertolini, E., Broomandkhoshbacht, N., Cooper, A., Culleton, B. J., Ferraz, T., Ferry, M., Furtwängler, A., Haak, W., Harkins, K., Harper, T. K., … Reich, D. (2018). Reconstructing the Deep Population History of Central and South America. *Cell*, *175*(5), 1185–1197.e22. https://doi.org/10.1016/j.cell.2018.10.027

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T. W., Jr, Orlando, L., Metspalu, E., Karmin, M., Tambets, K., Rootsi, S., Mägi, R., Campos, P. F., Balanovska, E., Balanovsky, O., Khusnutdinova, E., Litvinov, S., … Willerslev, E. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, *505*(7481), 87–91. https://doi.org/10.1038/nature12736

Rands, C. M., Meader, S., Ponting, C. P., & Lunter, G. (2014). 8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS Genetics*, *10*(7), e1004525. https://doi.org/10.1371/journal.pgen.1004525

Rasmussen, M., Anzick, S. L., Waters, M. R., Skoglund, P., DeGiorgio, M., Stafford, T. W., Jr, Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S. M., Poznik, G. D., Gudmundsdottir, V., Yadav, R., Malaspinas, A.-S., White, S. S., 5th, Allentoft, M. E., Cornejo, O. E., Tambets, K., Eriksson, A., … Willerslev, E. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature*, *506*(7487), 225–229.

https://doi.org/10.1038/nature13025

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J. S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., Bertalan, M., Nielsen, K., Gilbert, M. T. P., Wang, Y., Raghavan, M., Campos, P. F., Kamp, H. M., Wilson, A. S., Gledhill, A., … Willerslev, E. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*, *463*(7282), 757–762. https://doi.org/10.1038/nature08835

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M. V., Rojas, W., Duque, C., Mesa, N., García, L. F., Triana, O., Blair, S., Maestre, A., Dib, J. C., Bravi, C. M., Bailliet, G., Corach, D., Hünemeier, T., … Ruiz-Linares, A. (2012). Reconstructing Native American population history. *Nature*, *488*(7411), 370–374. https://doi.org/10.1038/nature11258

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*(7263), 489–494. https://doi.org/10.1038/nature08365

Saag, L., Varul, L., Scheib, C. L., Stenderup, J., Allentoft, M. E., Saag, L., Pagani, L., Reidla, M., Tambets, K., Metspalu, E., Kriiska, A., Willerslev, E., Kivisild, T., & Metspalu, M. (2017). Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe. *Current Biology: CB*, *27*(14), 2185–2193.e6. https://doi.org/10.1016/j.cub.2017.06.022

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, *239*(4839), 487–491. https://doi.org/10.1126/science.2448875

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site

analysis for diagnosis of sickle cell anemia. *Science*, *230*(4732), 1350–1354. https://doi.org/10.1126/science.2999980

Salmela, E., Lappalainen, T., Fransson, I., Andersen, P. M., Dahlman-Wright, K., Fiebig, A., Sistonen, P., Savontaus, M.-L., Schreiber, S., Kere, J., & Lahermo, P. (2008). Genome-wide analysis of single nucleotide polymorphisms uncovers population structure in Northern Europe. *PloS One*, *3*(10), e3519. https://doi.org/10.1371/journal.pone.0003519

Sánchez-Quinto, F., Malmström, H., Fraser, M., Girdland-Flink, L., Svensson, E. M., Simões, L. G., George, R., Hollfelder, N., Burenhult, G., Noble, G., Britton, K., Talamo, S., Curtis, N., Brzobohata, H., Sumberova, R., Götherström, A., Storå, J., & Jakobsson, M. (2019). Megalithic tombs in western and northern Neolithic Europe were linked to a kindred society. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(19), 9469–9474. https://doi.org/10.1073/pnas.1818037116

Scheib, C. L., Li, H., Desai, T., Link, V., Kendall, C., Dewar, G., Griffith, P. W., Mörseburg, A., Johnson, J. R., Potter, A., Kerr, S. L., Endicott, P., Lindo, J., Haber, M., Xue, Y., Tyler-Smith, C., Sandhu, M. S., Lorenz, J. G., Randall, T. D., … Kivisild, T. (2018). Ancient human parallel lineages within North America contributed to a coastal expansion. *Science*, *360*(6392), 1024–1027. https://doi.org/10.1126/science.aar6851

Schiffels, S., Haak, W., Paajanen, P., Llamas, B., Popescu, E., Loe, L., Clarke, R., Lyons, A., Mortimer, R., Sayer, D., Tyler-Smith, C., Cooper, A., & Durbin, R. (2016). Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nature Communications*, *7*, 10408. https://doi.org/10.1038/ncomms10408

Schuenemann, V. J., Kumar Lankapalli, A., Barquera, R., Nelson, E. A., Iraíz Hernández, D., Acuña Alonzo, V., Bos, K. I., Márquez Morfín, L., Herbig, A., & Krause, J. (2018).

Historic Treponema pallidum genomes from Colonial Mexico retrieved from archaeological remains. *PLoS Neglected Tropical Diseases*, *12*(6), e0006447. https://doi.org/10.1371/journal.pntd.0006447

Skoglund, P., Mallick, S., Bortolini, M. C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M. L., Salzano, F. M., Patterson, N., & Reich, D. (2015). Genetic evidence for two founding populations of the Americas. *Nature*, *525*(7567), 104–108. https://doi.org/10.1038/nature14895

Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M. T. P., Götherström, A., & Jakobsson, M. (2012). Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science*, *336*(6080), 466–469. https://doi.org/10.1126/science.1216304

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M. V., Miller, R., Stewart, J. R., Derevianko, A. P., Jacobs, Z., Li, B., Roberts, R. G., Shunkov, M. V., de Lumley, H., Perrenoud, C., Gušić, I., … Meyer, M. (2017). Neandertal and Denisovan DNA from Pleistocene sediments. *Science*, *356*(6338), 605–608. https://doi.org/10.1126/science.aam9695

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V., & Richards, M. B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *American Journal of Human Genetics*, *84*(6), 740–759. https://doi.org/10.1016/j.ajhg.2009.05.001

Stannard, D. E. (1993). *American Holocaust: The Conquest of the New World*. Oxford University Press, USA. https://play.google.com/store/books/details?id=RzFsODcGjfcC

Svizzero, S. (2017). *Persistent controversies about the neolithic revolution*. https://hal.univ-

reunion.fr/hal-02145483/document

Taylor, P. G. (1996). Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. *Molecular Biology and Evolution*, *13*(1), 283–285. https://doi.org/10.1093/oxfordjournals.molbev.a025566

Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., Kauffman, E., Bonné-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J. R., Mehdi, S. Q., Seielstad, M. T., Wells, R. S., Piazza, A., Davis, R. W., Feldman, M. W., … Oefner, P. J. (2000). Y chromosome sequence variation and the history of human populations. *Nature Genetics*, *26*(3), 358–361. https://doi.org/10.1038/81685

Vågene, Å. J., Herbig, A., Campana, M. G., Robles García, N. M., Warinner, C., Sabin, S., Spyrou, M. A., Andrades Valtueña, A., Huson, D., Tuross, N., Bos, K. I., & Krause, J. (2018). Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology & Evolution*, *2*(3), 520–528. https://doi.org/10.1038/s41559-017-0446-6

Velsko, I. M., Fellows Yates, J. A., Aron, F., Hagan, R. W., Frantz, L. A. F., Loe, L., Martinez, J. B. R., Chaves, E., Gosden, C., Larson, G., & Warinner, C. (2019). Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. *Microbiome*, *7*(1), 102. https://doi.org/10.1186/s40168-019-0717-3

Velsko, I., Skourtanioti, E., & Brandt, G. (n.d.). *Ancient DNA Extraction from Skeletal Material v1 (protocols.io.baksicwe)* [Data set]. https://doi.org/10.17504/protocols.io.baksicwe

Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R. Y., Fiddyment, S., Speller, C., Hendy, J., Charlton, S., Luder, H. U., Salazar-García, D. C., Eppler, E., Seiler, R., Hansen, L. H., Castruita, J. A. S., …

Cappellini, E. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics*, *46*(4), 336–344. https://doi.org/10.1038/ng.2906

Wei, W., Ayub, Q., Chen, Y., McCarthy, S., Hou, Y., Carbone, I., Xue, Y., & Tyler-Smith, C. (2013). A calibrated human Y-chromosomal phylogeny based on resequencing. *Genome Research*, *23*(2), 388–395. https://doi.org/10.1101/gr.143198.112

Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A. G., Alt, K. W., Caramelli, D., Dresely, V., Farrell, M., Farrer, A. G., Francken, M., Gully, N., Haak, W., Hardy, K., Harvati, K., Held, P., Holmes, E. C., … Cooper, A. (2017). Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature*, *544*(7650), 357–361. https://doi.org/10.1038/nature21674

Woodward, S. R., Weyand, N. J., & Bunnell, M. (1994). DNA sequence from Cretaceous period bone fragments. *Science*, *266*(5188), 1229–1232. https://doi.org/10.1126/science.7973705

World Population Clock: 7.8 Billion People (2020) - Worldometer. (n.d.). Retrieved July 2, 2020, from https://www.worldometers.info/world-population/

Wright, J. L., Wasef, S., Heupink, T. H., Westaway, M. C., Rasmussen, S., Pardoe, C., Fourmile, G. G., Young, M., Johnson, T., Slade, J., Kennedy, R., Winch, P., Pappin, M., Sr, Wales, T., Bates, W. B., Hamilton, S., Whyman, N., van Holst Pellekaan, S., McAllister, P. J., … Lambert, D. M. (2018). Ancient nuclear genomes enable repatriation of Indigenous human remains. *Science Advances*, *4*(12), eaau5064. https://doi.org/10.1126/sciadv.aau5064

9. Summary

The field of archaeogenetics began in 1984, with the first studies of ancient DNA analysing the genetic material of ancient Egyptian mummified remains of humans, and an extinct equine species respectively. A number of technological and technical advances have been instrumental in expanding this field, which is currently experiencing much interest from both researchers and the public. The analysis of ancient DNA from archaeological material has been used to gain insights into disease pandemics in the past, the population dynamics of human and animal populations, the oral microbiome of ancient populations, and more. In this thesis I focus on the insights that can be gained about human population history through archaeogenetic analysis of human ancient DNA.

Manuscript A presents the results from analysis of 14 ancient genomes from Fennoscandia and a present-day genome from a Saami individual (for whom sequencing data had previously been released), spanning a temporal range of 3,500 years. This includes the earliest sampled ancient population that carries Siberian ancestry in northeastern Europe as well as the first ancient nuclear genomes from Finland. We use a modelling approach to quantify the proportion of Siberian ancestry in present-day and ancient populations from the region. The results of this analysis suggest that Siberian ancestry entered the region between 5,000 and 3,500 yBP, although an earlier entry is possible if this ancestry remained geographically restricted to northern Fennoscandia. In the present-day, the presence of Siberian ancestry is most pronounced in populations speaking Uralic languages. However, while Siberian ancestry is present in the region already 3,500 yBP, that time predates most linguistic estimates for the spread of Uralic languages into Europe. Manuscript A therefore concludes that if Uralic languages spread into Europe alongside Siberian ancestry, then the extent of Siberian ancestry observed in present-day populations is the result of multiple waves of this ancestry in the area. Additionally, comparative analysis of genetic data retrieved from an Iron Age population in southern Finland revealed that this population was more closely related to the present-day Saami than to the Finnish population. This result confirms that the ancestors of the Saami inhabited a larger geographic range than the Saami population does today.

Manuscript B presents genomic data from 15 ancient Fuego-Patagonian individuals, covering a temporal range between 5,000 and 300 yBP. We reveal that ancient Fuego-Patagonians derive the majority of their ancestry from a lineage that was thought to have been replaced in the rest of the continent. Between 5,000 and 1,100 yBP, a genetic shift is observed in the Kaweskar population, with the arrival of a minor ancestry component that is maximised in a population that inhabited the California Channel Islands 5,200 yBP. Following the arrival of this component, genetic continuity is detected within Fuego-Patagonian populations. When compared to present-day Native American populations, the ancient Fuego-Patagonians show similar affinity to both Andean and Amazonian populations, with groups with discernable ancestry related to the Channel Islands appearing more closely related to Amazonian populations. We constructed a demographic model of Native American populations based on rare variation. We then inferred the most likely branching point for different ancient Native American individuals on this demographic model, based on the sharing of rare alleles between the ancient individuals and the populations in the model. We observe the inferred branching point of ancient Fuego-Patagonians to be closer to the split time between the different present-

day Native American populations, while the confidence interval around each placement included multiple Native American branches in most cases. These findings confirmed the broader Native American affinity exhibited by the ancient Fuego-Patagonians. We argue that this broad affinity is the result of population history events within South America. Finally, the most recent sampled individuals (~300 yBP) show no signs of admixture with European or African populations, a result that suggests Fuego-Patagonian populations remained isolated from European-contact for longer than native populations in the rest of the continent.

Manuscript C presents the interdisciplinary study of three individuals who were discovered in a mass burial at a colonial-time hospital in Mexico City, and who exhibited dental modifications consistent with certain African populations. Previously, this hospital was thought to have been exclusively for the treatment of indigenous people. Immunogenetic and uniparental markers confirmed the African origin of these individuals, while analysis of nuclear DNA from these individuals identified that all individuals were most closely related to present-day West African populations. However, each individual was most closely related to a different present-day population. Strontium ratios obtained from the molars of these individuals revealed that they arrived in Mexico after the 8th year of their lives, the time when these molars develop. Together the archaeogenetic and strontium analysis results point to these individuals being first-generation Africans, brought over to Mexico during colonial times. Osteological analyses found signatures of malnutrition, poor hygiene, hardship, and conflict on the remains of the individuals. A complete genome of *Treponema pallidum sub. pertenue* (causative agent of yaws) was recovered from one individual. Phylogenetic analysis of this genome revealed it is most closely related to treponemal genomes from present-day Ghana and colonial Mexico. Finally, the complete genome of a Hepatitis B Virus (HBV) was recovered from another individual. The recovered HBV genome was most closely related to HBV genotypes isolated in West Africa in the present. In concert, these results let us reconstruct the life history of these three individuals, and gain insights into the role of the trans-atlantic slave trade in the dissemination of disease.

This thesis explores the effects of admixture and migration on human population history, and the breadth of information that archaeogenetic approaches make available to researchers of human history.

10. Zusammenfassung

Das Fach Archäogenetik nahm 1984 seinen Ursprung mit ersten Studien an alter DNA (aDNA). Dabei wurde das genetische Material altägyptischer, mumifizierter Überreste von Menschen und einer ausgestorbenen Pferdeart analysiert. Eine Reihe technologischer Fortschritte haben dazu beigetragen, dieses Fachgebiet, das derzeit sowohl von Wissenschaftlern als auch in der Öffentlichkeit auf großes Interesse stößt, immer weiter auszubauen. Mittels Analyse alter DNA aus archäologischen Kontexten wurden unter anderem Erkenntnisse über Krankheitspandemien in der Vergangenheit, die Populationsdynamik menschlicher und tierischer Populationen, die Entwicklung des oralen Mikrobioms und vieles mehr gewonnen. In der vorliegenden Arbeit konzentriere ich mich auf archäogenetische Analysen zur menschlichen Populationsgeschichte.

Manuskript A stellt die Ergebnisse einer Analyse von 14 alten Genomen aus Fennoskandinavien und einem modernen Genom eines Individuums der Sámi vor. Die zeitliche Tiefe dieses Datensatzes erstreckt sich über einen Zeitraum von 3500 Jahren. Er umfasst die früheste beprobte Population sibirischer Abstammung in Nordosteuropa sowie die ersten alten Kerngenome aus Finnland. Wir implementierten einen Modellierungsansatz, um den Anteil sibirischer Abstammung (Ancestry) an den modernen und alten Populationen der Region zu quantifizieren. Die Ergebnisse dieser Analyse deuten darauf hin, dass die sibirische Abstammungskomponente zwischen 5000 und 3500 BP erstmals in die Region eingebracht wurde, obgleich ein früherer Eintrag möglich ist, sofern er zunächst geographisch auf Nordfennoskandiavien beschränkt blieb. Heute ist die Präsenz sibirischer Abstammung in uralisch-sprechenden Bevölkerungsgruppen am stärksten ausgeprägt. Das Aufkommen sibirischer Abstammung in der Region um 3500 BP geht jedoch den meisten linguistischen Schätzungen für die Verbreitung der uralischen Sprachen in Europa deutlich voraus. Manuskript A kommt daher zum Schluss, dass - sofern sich die uralischen Sprachen tatsächlich zusammen mit einer sibirischen Abstammungskomponente in Europa ausgebreitet haben - der heute messbare Anteil sibirischer Abstammung in den entsprechenden Bevölkerungsgruppen das Ergebnis mehrerer Ausbreitungswellen dieser Abstammung ist. Darüber hinaus ergab eine vergleichende Analyse des genetischen Profils einer eisenzeitlichen Population in Südfinnland, dass diese enger mit den heutigen Sámi als mit der heutigen finnischen Bevölkerung verwandt war. Dieses Ergebnis bestätigt, dass die Vorfahren der Sámi in einem größeren geographischen Gebiet lebten als die heutige Sámi-Population.

Manuskript B präsentiert aDNA Daten von 15 Individuen aus archäologischen Kontexten in Feuerland, die ein Zeitfenster zwischen 5000 und 300 BP abdecken. Wir konnten zeigen, dass diese Individuen einen Großteil ihrer Ancestry von einer Abstammungslinie beziehen, von der man annahm, dass sie überall im südamerikanischen Kontinent ersetzt wurde. Zwischen 5000 und 1100 BP lässt sich eine genetische Verschiebung in der Kawesqar-Population beobachtet: Die Ankunft einer kleineren Abstammungskomponente, die sich in heute verfügbaren Daten maximal in einer Bevölkerung wiederfindet, die 5200 BP auf den kalifornischen Channel Islands lebte. Nach der Ankunft dieser Komponente stabilisiert sich die feuerländische Population, was sich in genetischer Kontinuität ausdrückt. Im Vergleich zu heutigen Populationen amerikanischer Ureinwohner zeigen die alten Feuerländer eine Affinität sowohl zu Anden- als auch zu Amazonas-Populationen, wobei Gruppen mit erkennbarer

Channel Island Komponente enger mit den Amazonas-Populationen verwandt scheinen. Auf Grundlage seltener, genetischer Variationen haben wir ein demographisches Modell dieser Populationen erstellt. Daraus konnten wir den wahrscheinlichsten Verzweigungspunkt der Abstammungslinien hier analysierter Individuen ableiten - basierend auf Überlappungen seltener Allele zwischen den alten Individuen und Populationen im Modell. Der Verzweigungspunkt der alten Feuerländer liegt zeitlich relative nah am Verzweigungspunkt verschiedener heutiger Populationen aus Süd- und Nordamerika und das Konfidenzintervall um jede Individuen-Platzierung in diesem Modell erlaubt keine eindeutige Zuordnung zu einer Abstammungslinie bestimmter heutiger Populationen. Diese Ergebnisse bestätigen also eine breite Affinität der alten Feuerländer zu heutigen Populationen. Wir argumentieren, dass diese breite Affinität das Ergebnis bevölkerungsgeschichtlicher Ereignisse innerhalb Südamerikas sein muss (und nicht die Konsequenz zu niedriger Auflösung). Schließlich zeigen die jüngsten untersuchten Individuen (~300 BP) keine Anzeichen einer Vermischung mit europäischen oder afrikanischen Populationen, was darauf hindeutet, dass die feuerländischen Populationen länger von europäischen Kontakten isoliert blieben als die Ureinwohner im restlichen Kontinent.

Manuskript C ist eine interdisziplinäre Studie an drei Individuen, die in einer Massenbestattung nahe eines kolonialzeitlichen Krankenhauses in Mexiko-Stadt entdeckt wurden und die zahnmedizinische Veränderungen aufweisen, wie sie sich auch bei bestimmten afrikanischen Bevölkerungsgruppen finden lassen. Zuvor war angenommen worden, dass dieses Krankenhaus ausschließlich zur Behandlung amerikanischer Ureinwohner bestimmt war. Immun-genetische und uniparentale Marker bestätigten die afrikanische Herkunft dieser Individuen und die Analyse von Kern-DNA erlaubte eine weitere Eingrenzung auf Verwandtschaft mit heutigen westafrikanischen Bevölkerungsgruppen, allerdings auf drei verschiedene, ja nach Individuum. Strontiumisotopenverhältnisse, die aus den Molaren der Individuen gewonnen wurden, zeigten, dass sie nach dem 8. Lebensjahr, dem Zeitpunkt der Entwicklung dieser Molaren, nach Mexiko gekommen sein müssen. Zusammengenommen weisen die archäogenetischen und isotopenanalytischen Ergebnisse darauf hin, dass es sich bei diesen Individuen um Immigranten der ersten Generation handelt, die während der Kolonialzeit von Afrika nach Mexiko gebracht wurden. Eine osteologische Analyse der Überresten ergab Anzeichen von Unterernährung, schlechter Hygiene und Gewalteinwirkung. Bei einem Individuum wurde ein vollständiges Genom von Treponema pallidum sub. pertenue (Erreger der Frambösie) nachgewiesen. Die phylogenetische Analyse dieses Genoms ergab, dass es am engsten mit Treponema-Varianten aus dem heutigen Ghana und dem kolonialen Mexiko verwandt ist. In einem anderen Individuum wurde das komplette Genom eines Hepatitis-B-Virus (HBV) wiedergefunden. Dieses HBV-Genom ist am engsten mit den in Westafrika isolierten HBV-Genotypen der Gegenwart verwandt. Die interdisziplinären Ergebnisse erlauben die Rekonstruktion eines Teils der Lebensgeschichte dieser drei Individuen und ermöglichen Einblicke in die Rolle des transatlantischen Sklavenhandels bei der Verbreitung von Krankheiten.

Die vorliegende Arbeit untersucht die Auswirkungen von Vermischung (admixture) und Migration auf die menschliche Populationsgeschichte und zeigt die Bandbreite an Informationen, die archäogenetische Ansätze den anthropologischen Wissenschaften eröffnen.

11. Eigenständigkeitserklärung

Entsprechend §5 Abs. 4 der Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena, erkläre ich, dass mir die geltende Promotionsordnung der Fakultät bekannt ist. Ich bezeuge, dass ich die vorliegende Dissertation selbst angefertigt habe und keine Textabschnitte eines Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel, persönliche Mitteilungen sowie Quellen in meiner vorliegenden Arbeit angegeben habe. Zudem habe ich alle Personen, die mir bei der Auswahl und Auswertung sowie bei der Erstellung der Manuskripte unterstützt haben, in der Auflistung der Manuskripte und den entsprechenden Danksagungen namentlich erwähnt. Zudem versichere ich, dass ich die Hilfe eines Promotionsberaters nicht in Anspruch genommen haben und auch Dritten von mir keine unmittelbaren sowie mittelbaren geldwerte Leistungen für Arbeiten, die im Zusammenhang mit dieser Dissertation stehen, erhalten haben. Die vorliegende Promotion wurde zuvor weder für eine staatliche oder andere wissenschaftliche Prüfung eingereicht, also auch einer anderen Hochschule als Dissertation vorgelegt.

Jena, den 02.04.2020

Thiseas Christos Lamnidis

12. Curriculum Vitae

Department of Archaeogenetics
Max Planck Institute for the Science of Human History
Kahlaische Strasse 10
Jena, Germany 07745

Kronfeldstraße 21
Jena, Germany 07745
+49 (0) 3641 686-638
lamnidis@shh.mpg.de

EDUCATION

Max Planck Institute for the Science of Human History. Jena, Germany.
PhD will be awarded by Friedrich-Schiller-Universität Jena.
PhD Thesis: *Using ancient DNA to study human population interactions through time.*

Durham University. Durham, UK
M.Sc. Evolutionary Anthropology, Department of Anthropology, September 2014
Dissertation: *Investigating the Geographic range of the Denisovans.*

University of York. York, UK
B.Sc. (Hons) Genetics, Department of Biology, July 2013
Thesis: *Investigation of the pigmentation profile of the Denisova hominin by means of comparative genomics.*

Standardised Tests

SAT results (2010):	Reading: 590	Mathematics: 660	Writing: 620
SAT subject tests (2010):	Biology: 690	Mathematics: 710	Chemistry: 780
GRE results (2014):	Verbal: 157	Quantitative: 164	Analytical Writing: 5.0

Courses

Introduction to data analysis using R. October – November 2019
MPI-SHH. Jena, Germany.

Scientific writing class with Linda Vigilant. November 2018
MPI-EVA. Leipzig, Germany.

ArcGIS 1 Introduction to GIS 10.4. ESRI. March 2017
MPI-SHH. Jena, Germany.

PhD Course in analysis of genotyping and next generation sequencing data in medical and population genetics – without exam. August 2016
University of Copenhagen. Copenhagen, Denmark.

PUBLICATIONS

In Prep Nägele, K., Posth, C., Orbegozo, M. I., de Armas, Y. C., Godoy, S. T. H., Herrera, U. G., Colon, M. N., Sandoval-Velasco, M., Mylopotamitaki, D., Radzeviciute, R., Laffoon, J., Pestle, W., Ramos-Madrigal, J., **Lamnidis, T. C.**, Schaffer, W. C., Carr, R. S., Day, J. S., Antúnez, C. A., Rivero, A. R., … Schroeder, H. (n.d.). Genomic insights into the early peopling of the Caribbean. *In Prep.*

In Press Barquera*, R., **Lamnidis*, T. C.**, Lankapalli*, A. K., Kocher*, A., Hernández-Zaragoza*, D. I., Nelson, E. A., Zamora-Herrera, A. C., Ramallo, P., Bernal-Felipe, N., Immel, A., Acuña-Alonzo, V., Barbieri, C., Roberts, P., Herbig, A., Kühnert, D., Márquez-Morfín, L., & Krause., J. (n.d.). Origin and health status of first-generation Africans from early Colonial Mexico. *In Prep.*

In Press Furtwängler, A., Rohrlach, A. B., **Lamnidis, T. C.**, Papac, L., Neumann, G. U., Siebke, I., Reiter, E., Steuri, N., Hald, J., Denaire, A., Schnitzler, B., Wahl, J., Ramstein, M., Schuenemann, V. J., Stockhammer, P. W., Hafner, A., Lösch, S., Haak, W., Schiffels, S., & Krause, J. (n.d.). Ancient genomes reveal parallel societies in Late Neolithic Switzerland. *In Prep.*

2019 Flegontov, P., Altınışık, N. E., Changmai, P., Rohland, N., Mallick, S., Adamski, N., Bolnick, D. A., Broomandkhoshbacht, N., Candilio, F., Culleton, B. J., Flegontova, O., Friesen, T. M., Jeong, C., Harper, T. K., Keating, D., Kennett, D. J., Kim, A. M., **Lamnidis, T. C.**, Lawson, A. M., … Schiffels, S. (2019). Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. Nature, 570(7760), 236–240. https://doi.org/10.1038/s41586-019-1251-y

2018 **Lamnidis*, T. C.**, Majander*, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., Khartanovich, V., Balanovsky, O., Ongyerth, M., Weihmann, A., Sajantila, A., Kelso, J., Pääbo, S., Onkamo, P., Haak, W., Krause, J., & Schiffels, S. (2018). Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. Nature Communications, 9(1), 5018. https://doi.org/10.1038/s41467-018-07483-5

2018 Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., **Lamnidis, T. C.**, Rohland, N., Nägele, K., Adamski, N., Bertolini, E., Broomandkhoshbacht, N., Cooper, A., Culleton, B. J., Ferraz, T., Ferry, M., Furtwängler, A., Haak, W., Harkins, K., Harper, T. K., … Reich, D. (2018). Reconstructing the Deep Population History of Central and South America. *Cell*, *175*(5), 1185–1197.e22. https://doi.org/10.1016/j.cell.2018.10.027

*** Equal contributions**

GRANTS, AWARDS, AND HONORS

Conference Bursary 2017

Human Evolution: Fossils, Archaeology, Ancient and Modern Times conference.

CONFERENCE PRESENTATIONS

Poster	**Lamnidis T. C.** et al. Insights into South American population history, from ancient DNA from Tierra del Fuego. *Society for Molecular Biology and Evolution 2016.* Gold Coast, Queensland. Australia. 3rd-7th July 2016.
Poster	**Lamnidis T. C.** et al. Genetic data from ancient individuals from Levänluhta provide insights into the population history of Finland. *Society for Molecular Biology and Evolution 2017*. Austin, TX. USA. 2nd-6th July 2017.
Poster	**Lamnidis T. C.** et al. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Human Evolution: Fossils, Ancient and Modern Genomes.* Hinxton, UK. 20th-22nd November 2017.
Oral presentation	**Lamnidis T.C.** et al. The origin and spread of Siberian ancestry in north-eastern Europe. *LAG2 - The origin and expansions of Uralic speaking populations.* Jena, Germany. 29th November – 1st December 2017.
Poster	**Lamnidis T.C.** et al. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Society for Molecular Biology and Evolution 2018.* Yokohama, Japan. 8th-12th July 2018.

RESEARCH EXPERIENCE

Max Planck Institute for the Science of Human History Jena, Germany
PhD student 2015-present
Advisors: Stephan Schiffels, Johannes Krause

TEACHING EXPERIENCE

"York Students in Schools" placement York, UK
University of York. York, UK. 2012-2013
Bioinformatics related tutorial creation and curation. Jena, Germany
MPI-SHH. Jena, Germany. 2016-present
Bioinformatics training of junior students. Jena, Germany
MPI-SHH. Jena, Germany. 2017-present
"Bare Bones Bash" IMPRS workshop Jena, Germany
MPI-SHH. Jena, Germany. 19.11.2019
"Reproducible research and Data Management" workshop – Bash basics Jena, Germany
MPI-SHH. Jena, Germany. 30.01.2020

RELATED PROFESSIONAL EXPERIENCE

Languages

Greek:	Native
English:	Fluent
German:	Conversational

Programming

Python 3:	Fluent. Preferred language for standalone scripting.
Bash:	Fluent.
R:	Proficient. Preferred language for data manipulation and plotting.
Nextflow:	Novice.

Multiple code examples available at: https://github.com/TCLamnidis.

13. Acknowledgements

I would like to extend my gratitude to the following people:

First and foremost, to my scientific mentors Stephan Schiffels and Johannes Krause, for teaching me much of what I have learnt, and for being role models in both my scientific endeavours, and work-life balance.

To Clemens Schmid, Stephan Schiffels, and Sabine Ziegler, for their help with the translation of the Summary and main findings of my work into German.

To my colleagues, for always greeting me with a friendly smile at the long queue to the coffee machine before a meeting. For listening intently to my ramblings during lab seminars and group meetings, and all the valuable feedback they have provided me over the years.

To my friends, in Jena and abroad, for the mental support you have offered me over the years. Your companionship was instrumental in my work-life balance, and I sincerely hope you keep providing me with it.

To Kerttu Majander, Karen Giffin, Aditya Kumar Lankpalli, Susanna Sabin, and Marieke van de Loosdrecht, for all the fruitful philosophical discussions, movie nights and pondering as to the meaning of life. Much of the colour of my life in Jena was coloured in by you lot. Blessed be the fruit.

To Rodrigo Barquera, for always being there at a moment's notice to grab McD's and/or a beer, and joke life's worries away.

To James Fellows Yates, my ex-officemate, partner-in-crime, rubber duck, mental support companion, and spirit animal. To Aida Andrades Valtueña, who put up with James and me, who is an awesome travel companion, but who cannot break good news in a believable way. To Maia Fellows Yates Andrades Valtueña, for making me smile whenever we meet, and blowing my mind with how fast she is growing up.

To Zandra Fagernäs, for her patience with me, for always trying to organise the chaos, always being happy to talk science with me, and teaching me that there's more to humans than… humans. Her support, especially in the weeks leading up to the submission of this thesis, has been amazing.

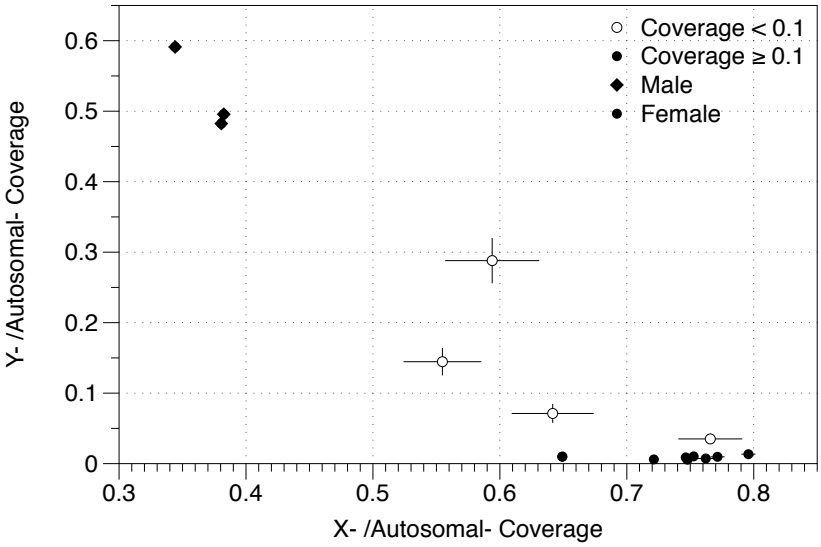
To my parents, Alexandros and Electra, for cheering me along my quest for knowledge and the sacrifices they made for me and my brother along the way. To my brother, Iason, who never fails to cheer me up, makes sure I don't take myself too seriously, and is always there to bring me back to reality. Most of all, I thank my family for never asking me when I am getting "a real job" and believing in me to turn my passion into a career.

14. Supplementary information

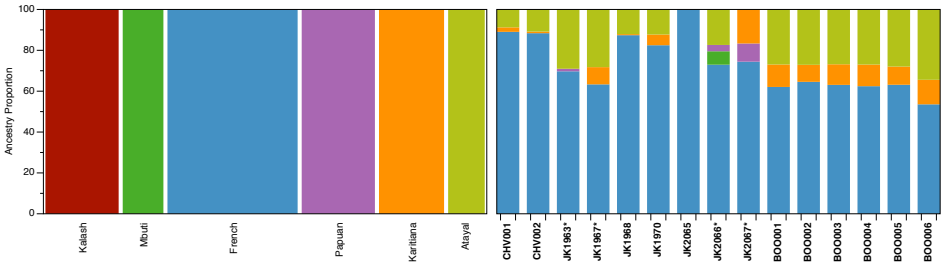
14.1 Supplementary Information for Manuscript A

Supplementary Material for Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe

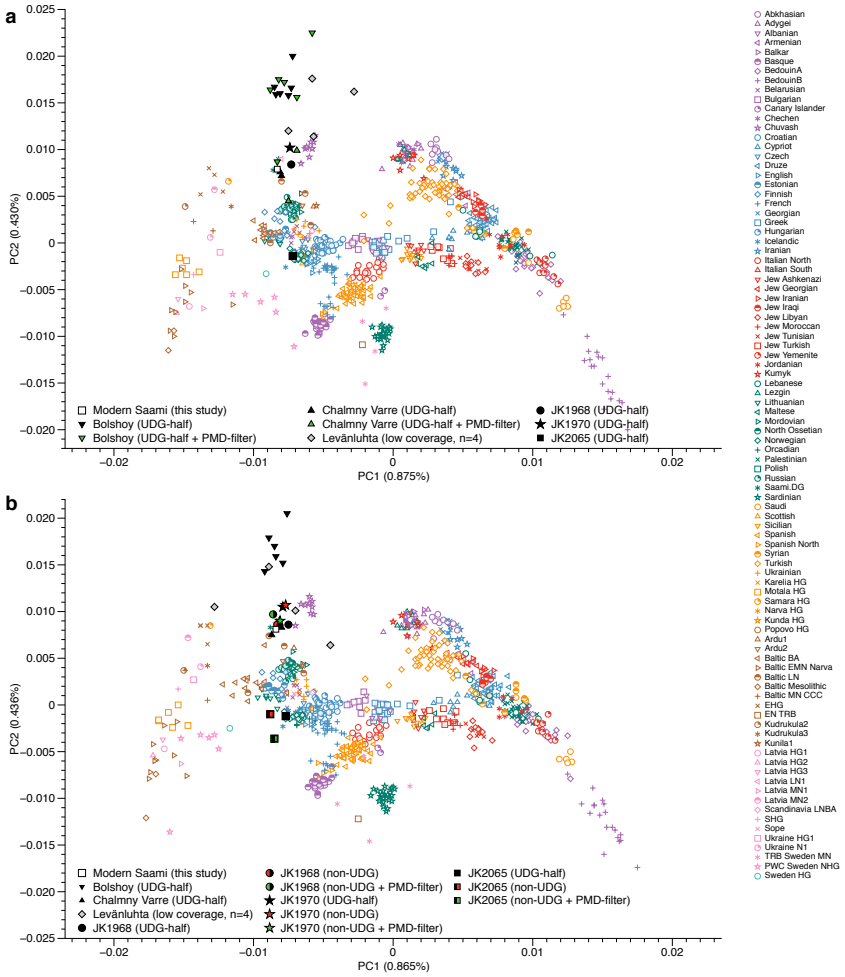
Supplementary Figures



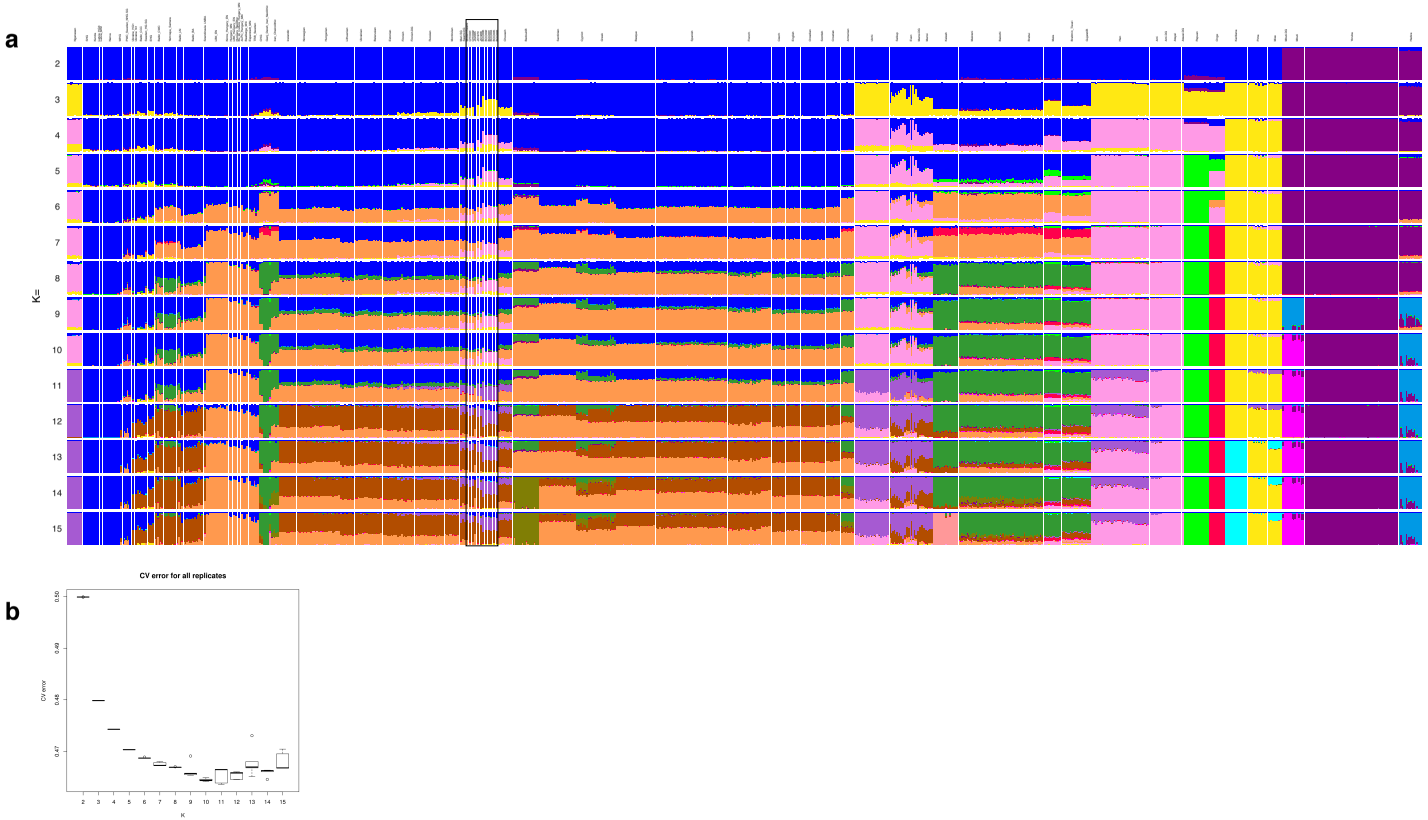
Supplementary Figure 1. Sex Determination. X-chromosomal coverage vs. Y-chromosomal coverage, normalised by autosomal coverage. Low autosomal coverage individuals (n = 4, excluded from downstream analyses) are shown as empty circles, and higher coverage individuals (n = 11) as filled circles. Error bars represent the uncertainty in the calculation of relative coverages, assuming reads map on the capture targets randomly (see Supplementary Text 2 for details). Source data are provided as a Source Data file.



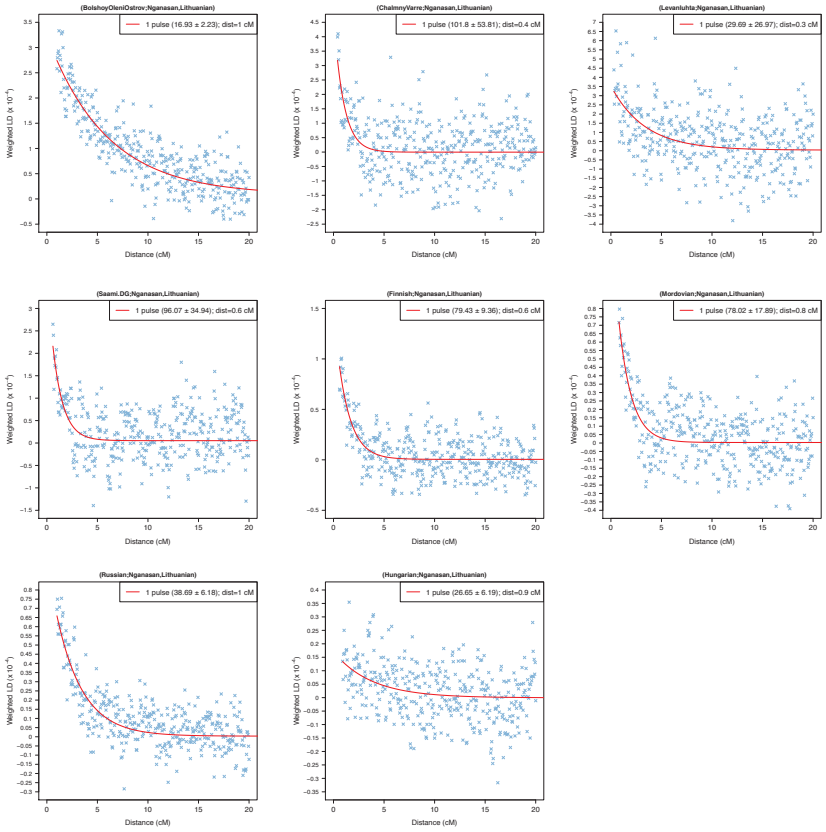
Supplementary Figure 2. Supervised ADMIXTURE. Individuals excluded from further analyses due to low coverage are signified with an asterisk. Among higher coverage genomes, the results within each population are homogeneous, with the exception of the outlier JK2065 within Levänluhta. Source data are provided as a Source Data file.



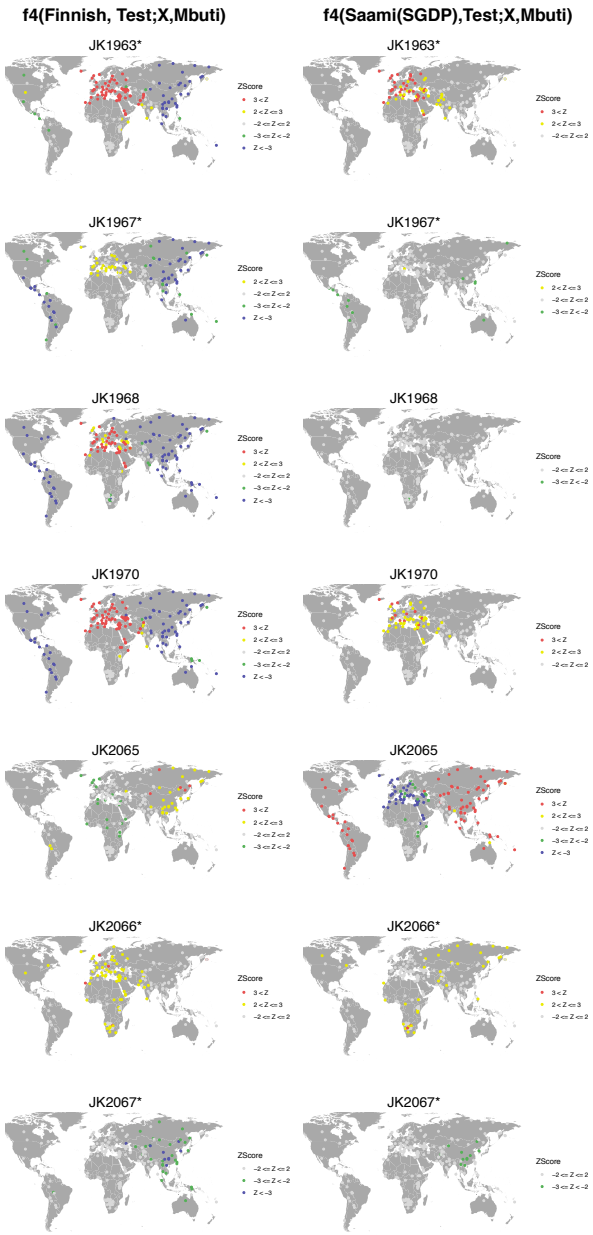
Supplementary Figure 3. Comparison of projection before and after PMD-filtering in PCA space. **a** PCA plot of Europe with individuals from this study projected on principal components constructed on the modern populations in the legend, using the option "shrinkmode: YES". For each individual from Bolshoy and Chalmny Varre, an additional projection is shown for the PMD-filtered dataset. Ancient individuals with fewer than 15,000 covered SNPs are shown in grey. **b** PCA plot of Europe with individuals from this study projected on principal components constructed using only transversion SNPs of the modern populations in the legend, using the option "shrinkmode: YES". For the three Levänluhta individuals with more than 15,000 covered SNPs, additional point for the PMD-filtered and the non-filtered datasets of the non-UDG treated libraries is shown. Ancient individuals with fewer than 15,000 covered SNPs are shown in grey. Source data are provided as a Source Data file.



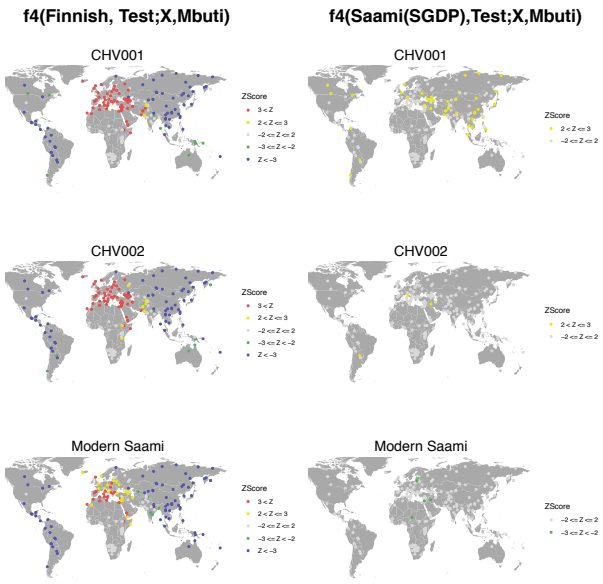
Supplementary Figure 4. ADMIXTURE analysis. **a** ADMIXTURE results. **b** CV Errors for 5 replicates per K value. Source data are provided as a Source Data file.



Supplementary Figure 5. Linkage Disequilibrium decay curves for ancient and modern populations. Minimum distance in cM used is the lowest distance for which ALDER provided results. Source data are provided as a Source Data file.



Supplementary Figure 6. Testing Levänluhta for cladality with modern Finns and/or Saami. f_4(Finnish, Test; X, Mbuti) & f_4(Saami, Test; X, Mbuti) comparison for multiple worldwide populations *X*, using ancient individuals from Levänluhta as *Test*. Asterisks denote individuals with low coverage (<15000 SNPs covered). Points are coloured based on bins of Z Score values, with warmer colours indicating higher affinity in Finns/Saami than in the *Test*, and vice versa for colder colours. Grey points indicate similar affinities toward population *X*. Map generated using ggplot2 and data from the Natural Earth project. Source data are provided as a Source Data file.



Supplementary Figure 7. Testing historical and modern Saami for cladality with modern Finns and/or Saami. f_4(Finnish, Test; X, Mbuti) & f_4(Saami, Test; X, Mbuti) comparison for multiple worldwide populations *X*, using ancient individuals from Chalmny Varre and the modern Saami genome from this study as *Test*. Points are coloured based on bins of Z Score values, with warmer colours indicating higher affinity in Finns/Saami than in the *Test*, and vice versa for colder colours. Grey points indicate similar affinities toward population *X*. Map generated using ggplot2 and data from the Natural Earth project. Source data are provided as a Source Data file.

Supplementary Tables

Supplementary Table 1. Deamination damage per library.

Library ID	UDG treatment	Damage 5' 1st base	Damage 5' 2nd base	Damage 3' 2nd base	Damage 3' 1st base
CHV001	Half	0.0287	0.0074	0.0121	0.031
CHV002	Half	0.0189	0.0062	0.0111	0.0244
JK1963	Half	0.0259	0.0071	0.0111	0.0294
JK1967	Half	0.0104	0.0048	0.009	0.0151
JK1968	Half	0.0284	0.0054	0.0088	0.0282
JK1968.nonUDG	None	0.094	0.0575	0.0521	0.0865
JK1970	Half	0.0239	0.0069	0.0109	0.029
JK1970.nonUDG	None	0.0953	0.0745	0.073	0.0937
JK2065	Half	0.0123	0.0045	0.0069	0.0155
JK2065.nonUDG	None	0.0471	0.0452	0.0448	0.0487
JK2066	Half	0.0146	0.0041	0.0059	0.0171
JK2067	Half	0.0104	0.004	0.0065	0.0129
BOO001	Half	0.0283	0.0075	0.0107	0.0302
BOO002	Half	0.0402	0.0078	0.0113	0.0333
BOO003	Half	0.0411	0.0073	0.0108	0.0398
BOO004	Half	0.0382	0.0085	0.0117	0.0344
BOO005	Half	0.045	0.0066	0.0108	0.0421
BOO006	Half	0.0381	0.0063	0.0094	0.0302

Supplementary Note 1

Extended Archaeological Information

Levänluhta

The Levänluhta site is located in the Isokyrö municipality at the southern Ostrobothnia region of Western Finland. The site represents a rarely observed case of lake burials, and is one of the most studied archaeological sites in Finland. Levänluhta context has been dated to the Iron Age in Finland (300-800 CE)[1,2] via some prestige artifacts assumed to have served as grave goods. The skeletal remains are, while numerous and well preserved, also anatomically disarticulated due to the gradual transition of the original lake environment to a marshland, and subsequent ditching and ploughing of the soil for agricultural use over the centuries. The remains of approximately 100 individuals are recognized from the cemetery to date.
The archaeological excavations were carried out by Oscar Rancken in 1886, A.M. Tallgren and Alfred Hackman from 1912 to 1913, Aarni Erä-Esko from the National Board of Antiquities from 1982 to 1984[1,2], followed by an archaeological survey of both Levänluhta and the immediate area around it in 2014. A comprehensive osteological analysis was reported by Tarja Formisto in 1993. The human remains are under the care of National Board of Antiquities, and stored currently at the National Museum of Finland. Skeletal element IDs for the analysed samples are: 2:1:a29 -L21 (JK1963), “milk tooth” -L47 (JK1967), 2:1:a16 -L20 (JK1968), 477 -L46 (JK1970), 2:1:a3 -L17 (JK2065), KM21814:735 -L22 (JK2066), 2:1:a2 -L16 (JK2067).

Chalmny Varre

The Chalmny-Varre Saami cemetery, associated with the two seasonal settlements of nomadic Kamensk Saami in the 18th century, is located on a small island in the middle flow of Ponoy River (center of Kola Peninsula). The burials have characteristics of Christian graves, combined with old traditional Saami rituals, such as birch cork pieces placed in the graves and masks (lichiny) of deceased carved on wooden crosses. Archaeological dating is confirmed by artifact findings from the graves and information on the wooden crosses which are marking the graves. The excavation, including an anthropological investigation, was organised by the Institute of Ethnography of N.N. Miklukho-Maclay Academy of Science of the USSR in the year 1976. Skeletal IDs for analysed samples are: CHV30 (CHV001), CHV38 (CHV002).

Bolshoy Oleniy Ostrov

Bolshoy Oleniy Ostrov (Great Reindeer Island), situated in the Kola Bay of the Barents Sea and separated from the mainland by Yekarerininsky Island and two straits, harbors the ancient cemetery of an unknown Early Metal Age culture. The preservation of artifacts made from bone and antler, wooden structures, as well as human remains is remarkable for the location and age this site represents. Altogether 19 skeletons of adults and children have been recognized from both single and collective burials of the site, together with more than 250 artifacts. Archaeological

surveys and excavations at the location were performed by G.D. Richter and S.F. Yegorov in 1925, by A.V. Schmidt on the USSR Academy of Science Kola Expedition in 1928, by N.N. Gurina in 1947–1948, as a part of Kola Expedition from the Leningrad Department of the Institute of Archaeology of the Academy of Sciences, and by V.Y. Shumkin from the same institute, later named as the Institute for the History of Material Culture Russian Academy of Sciences (RAS), starting from 1998-1999 and continuing in 2001-1004. Apart from these excavations, approximately 25 burials were revealed in 1934 during the construction of fortifications. Four finds are known to have been stored by the USSR Academy of Sciences at the time, but the location of all other remains from this instance is unknown. Part of the cemetery was never excavated and has possibly been destroyed by erosion. Morphological analyses, largely concentrated on cranial characteristics, have been performed by S.D. Sinitsyn in 1930, and V.P. Yakmov in 1953 and V.G. Moiseyev and V.I. Khartanovich in 2012[3]. A radiocarbon date for the site is provided by Moiseyev and Khartanovich as 3473±42 BP [3], which we calibrated as described in Methods. We note that radiocarbon dating on individuals with marine-based diets can be affected by the marine reservoir effect, which may result in overestimates of the true date[4]. Other dates provided by Murashkin et al (2016)[5] also date the site to the middle to late 2nd millennium BC. The human remains are stored at Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS in St. Petersburg, with the exception of burial 13, the remains of which are admitted to the Historical Museum in Polarnyi, Murmansk Oblast.[3] Skeleteal IDs for analysed samples are: BOO57.1 (BOO001), BOO72.1 (BOO002), BOO72.4 (BOO003), BOO72.7 (BOO004), BOO72.10 (BOO005), BOO72.15 (BOO006).

Supplementary Note 2

Assessment of power to detect contamination using continental supervised ADMIXTURE analysis.

To test the power of our approach to identify contamination using supervised ADMIXTURE using a set of continental populations as defined clusters, we devised a way to contaminate our data in silico. We assume that the rate of nuclear contamination in a library will be directly related to the proportion of reads mapping to the nuclear chromosomes that are of contaminant origin. Therefore, when a random draw genotyping approach is used, that contamination rate should also be directly proportional to the rate of genotypes called from the reads of the contaminant source. This assumption is violated if the distribution of contaminant and ancient DNA fragments across the nuclear genome is not uniform.

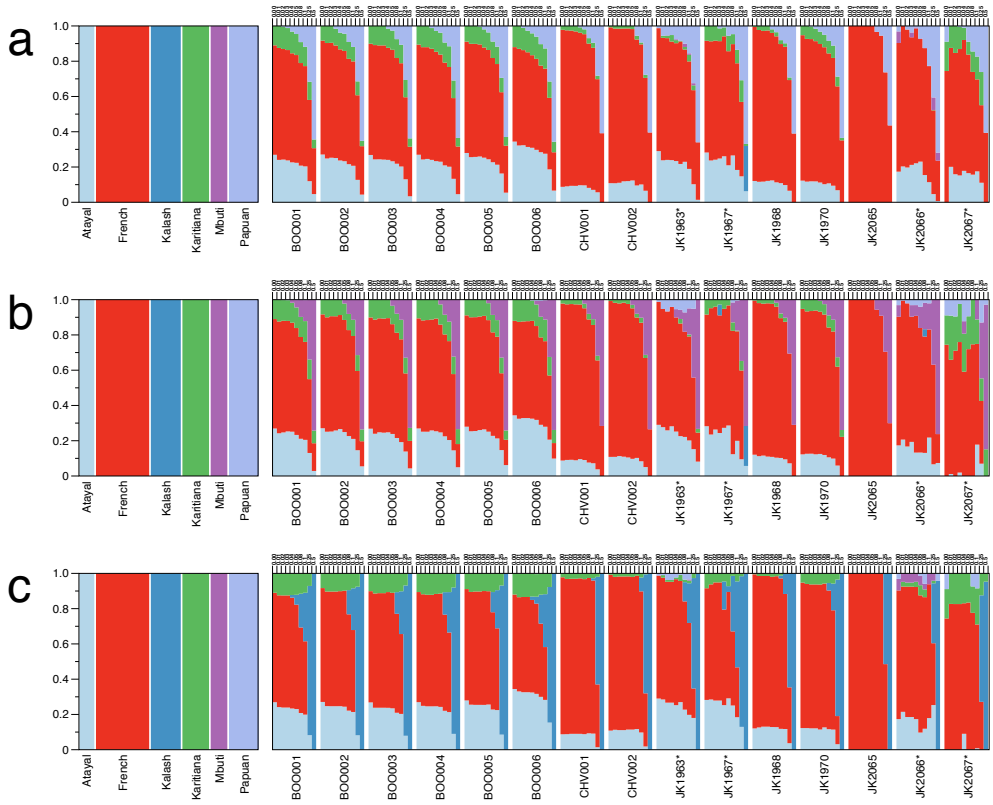
As a means to contaminate genotypes in silico, we provide ContaminateGenotypes.py (https://github.com/TCLamnidis/ContaminateGenotypes), a python script that contaminates the genotypes of a set of Sample individuals with genotypes from a provided contaminant individual, at multiple specified rates. We used this script to create a set of dummy individuals contaminated by one of seven contaminant individuals at nine different contamination rates.

As contamination sources we provided one individual from each population used as a pre-defined continental cluster during the supervised ADMIXTURE analysis (Atayal: NA13597, French: HGDP00511, Kalash: HGDP00267, Karitiana: HGDP00995, Mbuti: HGDP00449, Papuan: HGDP00540). We additionally used one Han Chinese individual (HGDP00774) as a contaminant to assess the power of this analysis to identify contamination that is not within the pre-defined continental clusters.

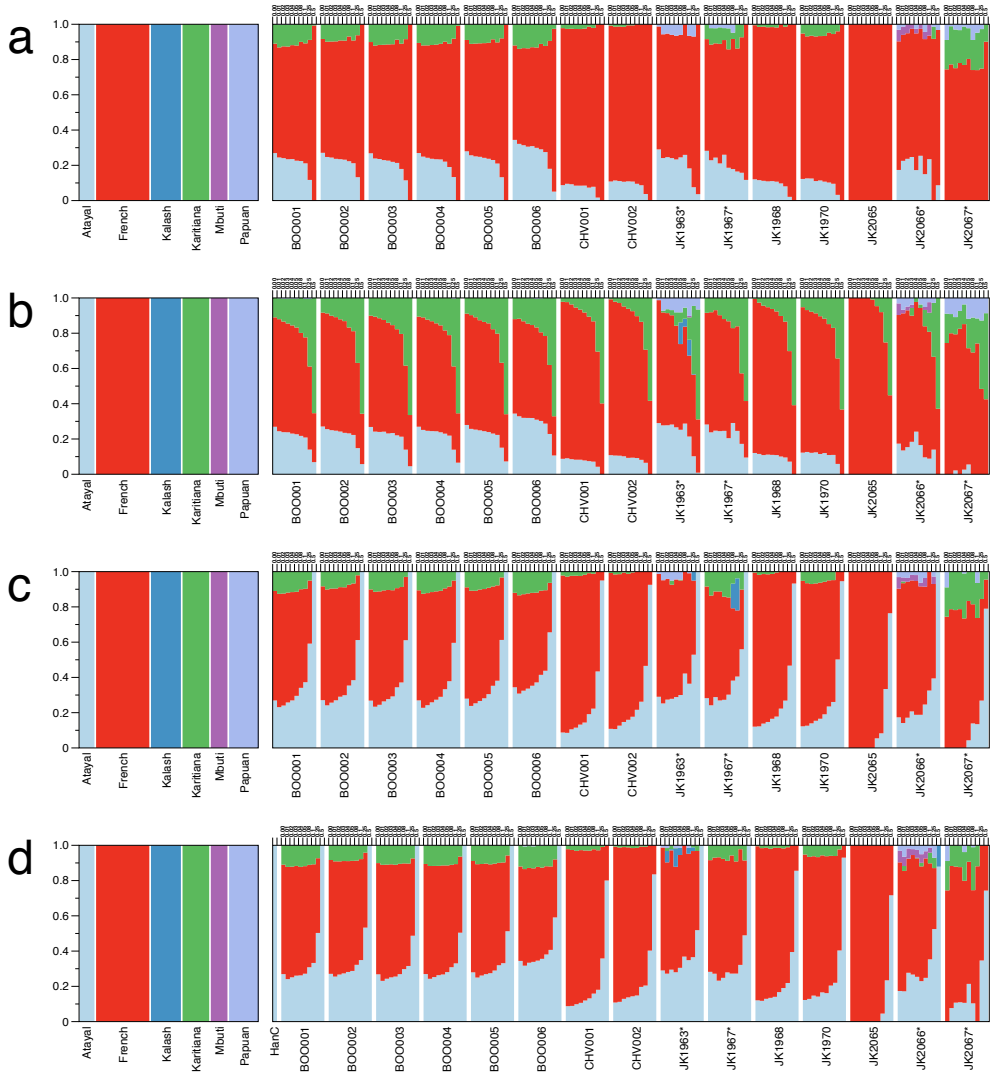
We generated dummy individuals at contamination rates of 0.01, 0.02, 0.03, 0.04, 0.05, 0.08, 0.10, 0.25, 0.50 for each of the contaminant sources. The dummy individuals were then separated by contaminant source and contamination rate, before running supervised admixture on each subset separately, to avoid artefacts in ADMIXTURE from essentially clonal individuals.

Our results show that contaminants that are distantly related to the ancestries within the ancient individuals (e.g. Mbuti and Papuan in our case), are detected from contamination rates of 0.05 and above, while more closely related ancestries, though ones not present in the dataset (e.g. Kalash) can be detected less consistently at that threshold, but quite consistently at contamination rates of 0.08 and above (Supplementary Figure 8).

Finally, contamination from sources whose continental ancestry was already detected in the uncontaminated data (e.g. Atayal, French, Karitiana) cannot be reliably detected as such (Supplementary Figure 9). We conclude that such an approach only has power to detect contamination from distantly related populations.



Supplementary Figure 8. Compiled supervised ADMIXTURE runs. Each individual from this study was contaminated at different rate by a distantly related contaminant individual. Low coverage (>15,000 SNPs) individuals marked with an asterisk. **a** Contaminant was Papuan. **b** Contaminant was Mbuti. **c** Contaminant was Kalash. Source data are provided as a Source Data file.



Supplementary Figure 9. Compiled supervised ADMIXTURE runs. Each individual from this study was contaminated at different rate by a contaminant individual from a genetic cluster that was already present in the uncontaminated data. Low coverage (>15,000 SNPs) individuals marked with an asterisk. **a** Contaminant was French. **b** Contaminant was Karitiana. **c** Contaminant was Atayal. **d** Contaminant was Han Chinese. Contaminant individual is labelled "HanC". Source data are provided as a Source Data file.

Supplementary Note 3

Sex determination with error bar calculation

We provide a python script (https://github.com/TCLamnidis/Sex.DetERRmine.git) that calculates the relative coverage of X and Y chromosomes, and their associated error bars, from the depth of coverage at specified SNPs. The error calculation relies on the assumption that reads are randomly and independently distributed across a list of specified SNPs. This is justified if SNPs are further apart from each other than typical sequencing length, which is largely the case for the SNP panel considered here (1240K). The script takes as input the output of the samtools depth command, which reports the depth at each of the specified SNPs. The depths are partitioned into three bins: Autosomal, X- and Y-chromosome, so that:

$$N = \sum_i N_i$$

Where N_i is the number of sequenced reads overlapping the SNPs within each bin i, and N is the total number of reads overlapping with SNPs within our capture panel.
We can then calculate the proportion of all sequenced reads that cover SNPs in bin i (p_i). Its associated error is then calculated as the error of the binomial distribution of N_i.

$$p_i = \frac{N_i}{N}$$
$$\mathrm{Err}(N_i) = \sqrt{N p_i (1 - p_i)}$$

The average SNP depth within bin i (d_i) can be calculated as the ratio of N_i to the number of SNPs within bin i (S_i). The error of the average SNP depth is proportional to the error associated with N_i.

$$d_i = \frac{N_i}{S_i}$$
$$\mathrm{Err}(d_i) = \frac{\mathrm{Err}(N_i)}{S_i}$$

The relative coverage on the X and Y chromosomes can then be calculated as the ratio of the average SNP depth on the X/Y chromosome over the average SNP depth on the autosomes. The error of this measurement can then be calculated using standard error propagation.

$$\mathrm{rate}_{X/Y} = \frac{d_{X/Y}}{d_{aut}}$$
$$\mathrm{Err}(\mathrm{rate}_{X/Y}) = \sqrt{\left(\mathrm{Err}(d_{X/Y}) \times \frac{1}{d_{aut}}\right)^2 + \left(\mathrm{Err}(d_{aut}) \times \frac{d_{X/Y}}{{d_{aut}}^2}\right)^2}$$

Supplementary References

1. Wessman, A. Levänluhta. A place of punishment, sacrifice or just a common cemetery? *Fennoscandia archaeologica* **XXVI,** 81–105 (2009).
2. Wessman, A. *et al.* Hidden and Remote: New Perspectives on the People in the Levänluhta Water Burial, Western Finland (c.ad 300–800). *European Journal of Archaeology* **21,** 431–454 (2018).
3. Moiseyev, V. G. & Khartanovich, V. I. Early Metal Age crania from Bolshoy Oleniy Island, Barents Sea. *Archaeology, Ethnology and Anthropology of Eurasia* **40,** 145–154 (2012).
4. Ascough, P., Cook, G. & Dugmore, A. Methodological approaches to determining the marine radiocarbon reservoir effect. *Progress in Physical Geography: Earth and Environment* **29,** 532–547 (2005).
5. Murashkin, A. I., Kolpakov, E. M., Shumkin, V. Y., Khartanovich, V. I. & Moiseyev, V. G. Kola Oleneostrovskiy grave field: a unique burial site in the European Arctic. *New Sites, New Methods; Iskos 21* (2016).

14.2 Supplementary Information for Manuscript C

Current Biology, Volume *30*

Supplemental Information

Origin and Health Status of First-Generation Africans from Early Colonial Mexico

Rodrigo Barquera, Thiseas C. Lamnidis, Aditya Kumar Lankapalli, Arthur Kocher, Diana I. Hernández-Zaragoza, Elizabeth A. Nelson, Adriana C. Zamora-Herrera, Patxi Ramallo, Natalia Bernal-Felipe, Alexander Immel, Kirsten Bos, Víctor Acuña-Alonzo, Chiara Barbieri, Patrick Roberts, Alexander Herbig, Denise Kühnert, Lourdes Márquez-Morfín, and Johannes Krause

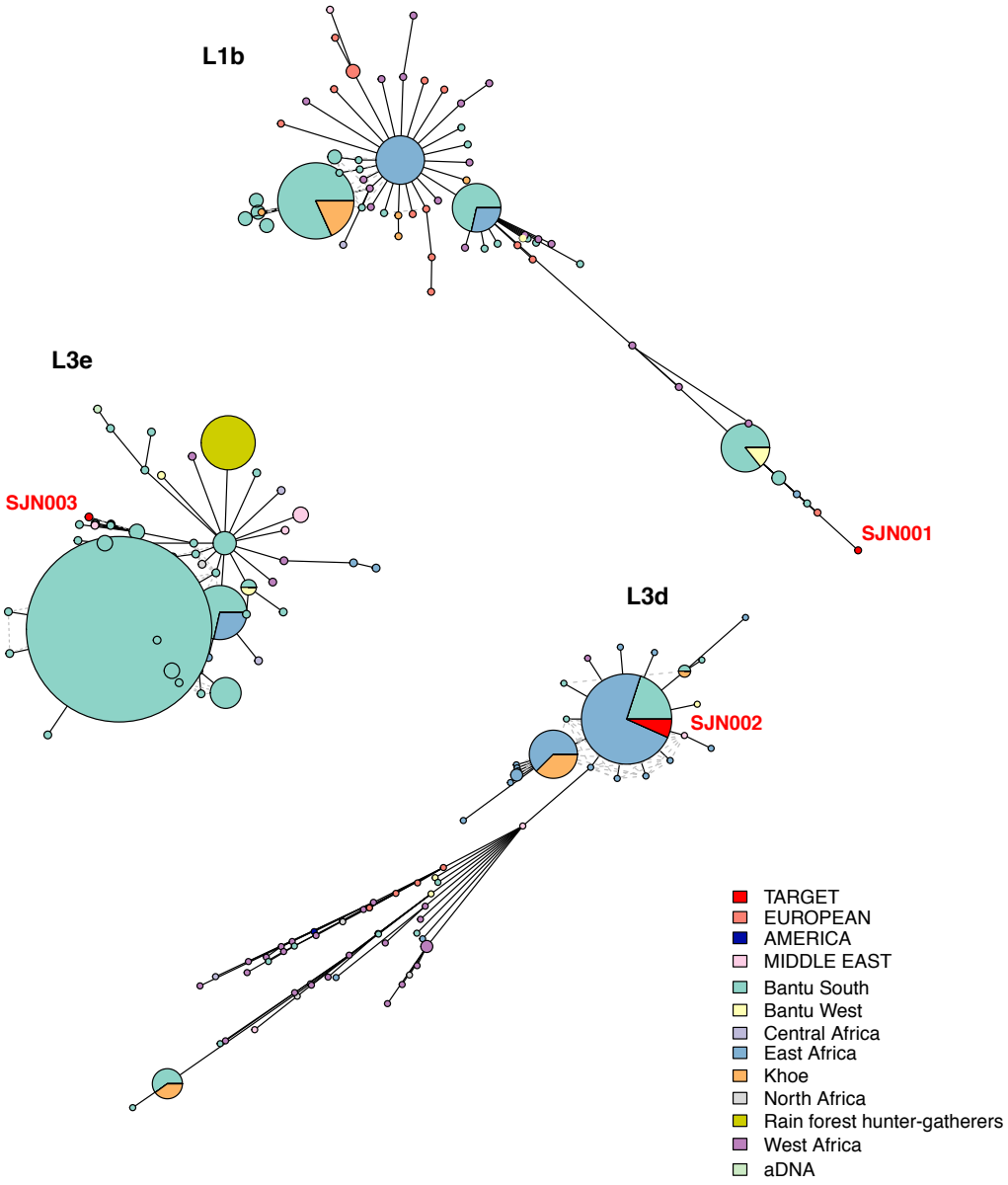


Figure S1. Phylogenetic networks built with mtDNA haplotypes for haplogroups L1b, L3d and L3e, related to STAR Methods, Uniparental markers. For each target sample, the 100 sequences with the closest F_{ST} distance are displayed. The comparative database includes sequences from non-African regions (in capital letters in the colour legend), Bantu South (Bantu speakers from Angola, Namibia, South Africa, Botswana, Zambia, Zimbabwe and Mozambique), Bantu West (Bantu speakers from Gabon), Central Africa (Chad and Uganda), East Africa (Comoros Island, Ethiopia, Kenya, Somalia, Tanzania), Khoe speakers (Angola, Botswana, Namibia, South Africa), North Africa (Egypt, Sudan, Algeria, Libya, Tunisia, Morocco), rainforest hunter-gatherers (Central Africa Republic, Gabon, Democratic Republic of Congo, Cameroon), and West Africa (Burkina Faso, Guinea-Bissau, Niger, Nigeria, Sao Tome, Senegal). aDNA refers to three ancient African genomes from the Stone Age and four from the Iron Age [S1-S16].

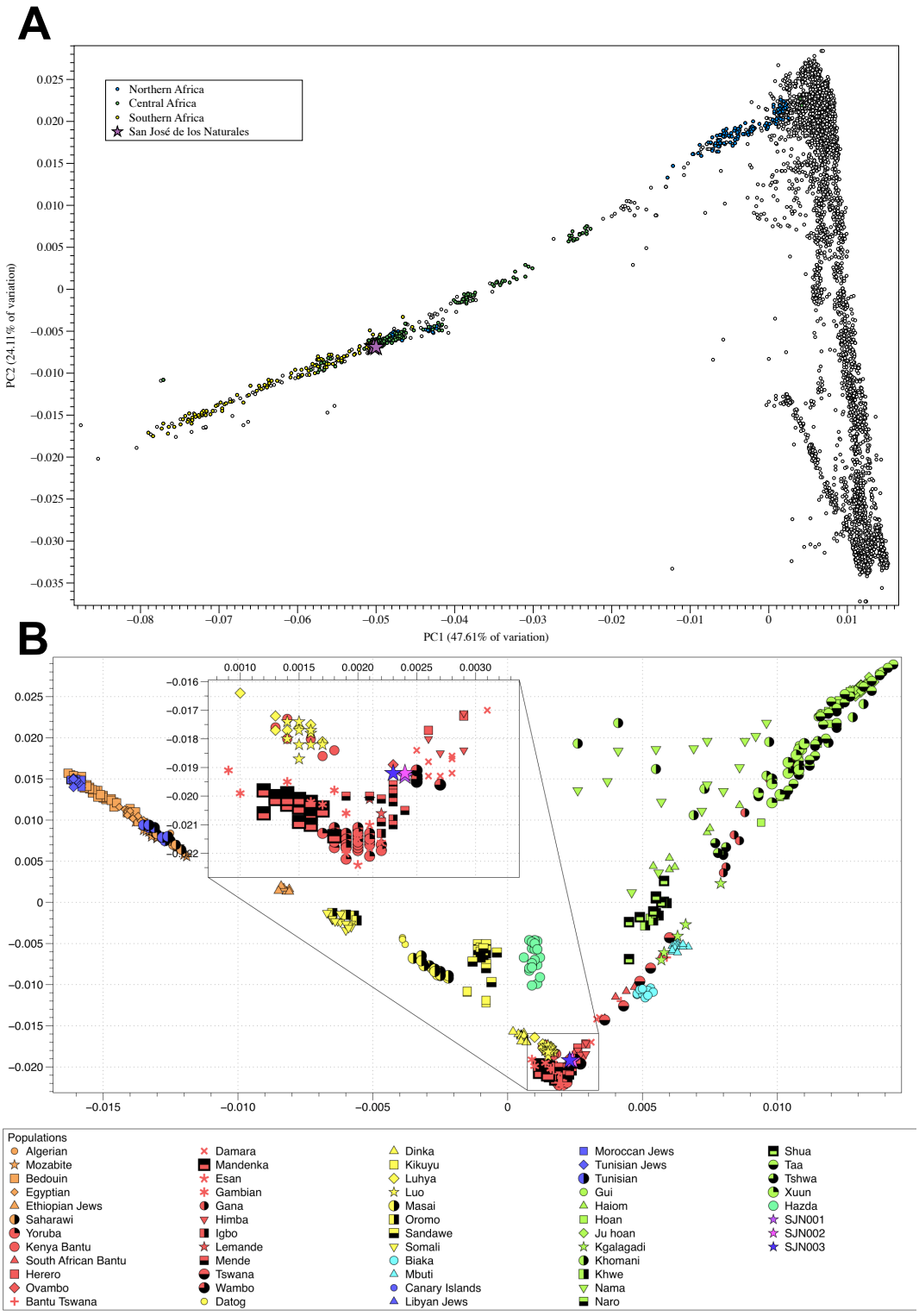


Figure S2. A. PCA (PC1 vs PC2) with worldwide populations showing the genetic relationship of SJN individuals to Africans, related to STAR Methods, Principal components analysis. B. PCA (PC1 vs PC2) showing the genetic relationship of SJN Africans to Western Africa Niger-Congo linguistic speakers, related to Results, Characterization of African ancestry; STAR Methods, Principal components analysis. PCA (PC1 vs PC2) showing the genetic relationship of SJN individuals (purple, pink and violet stars) to Western Africa Niger-Congo linguistic speakers.

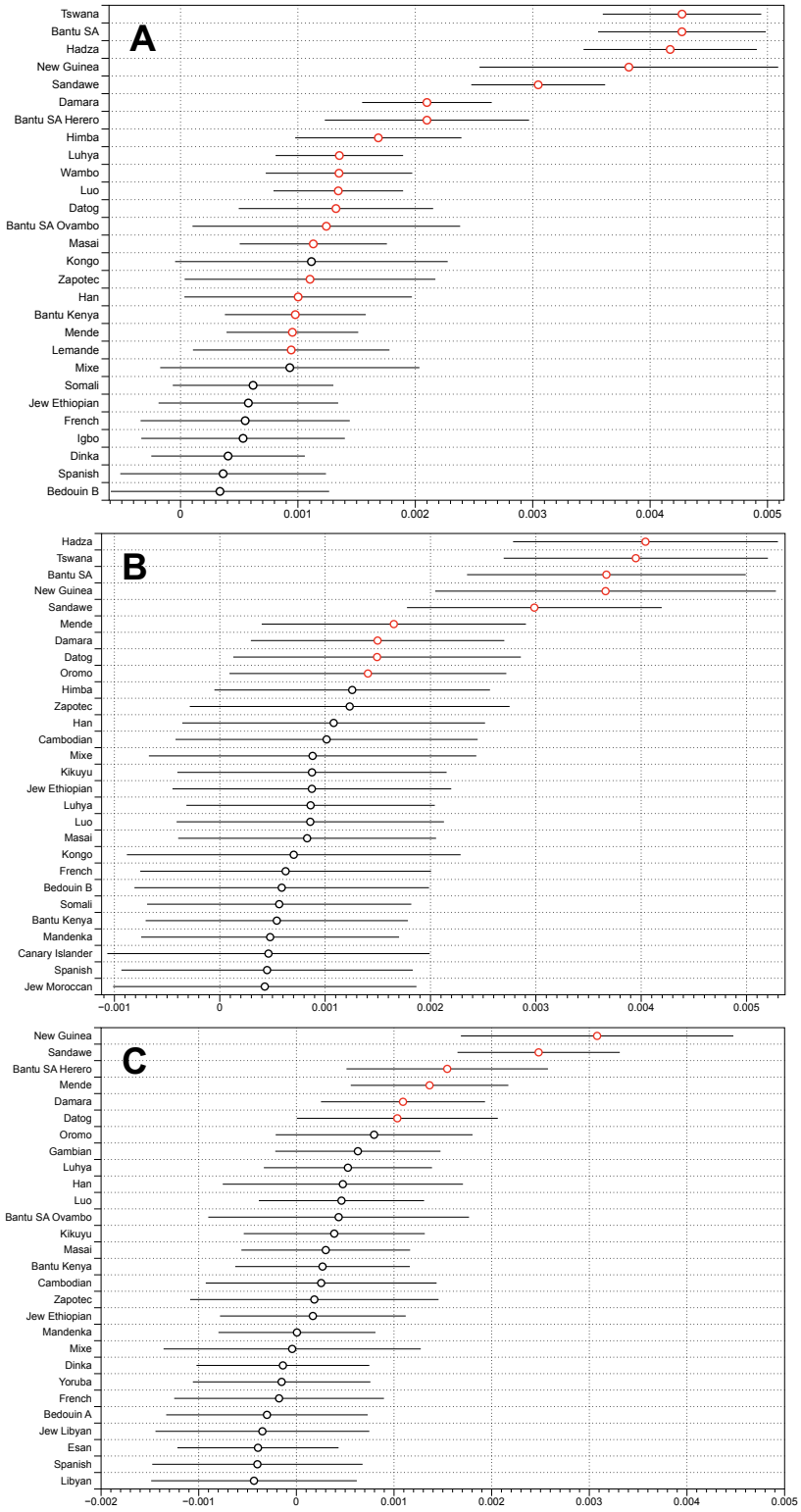
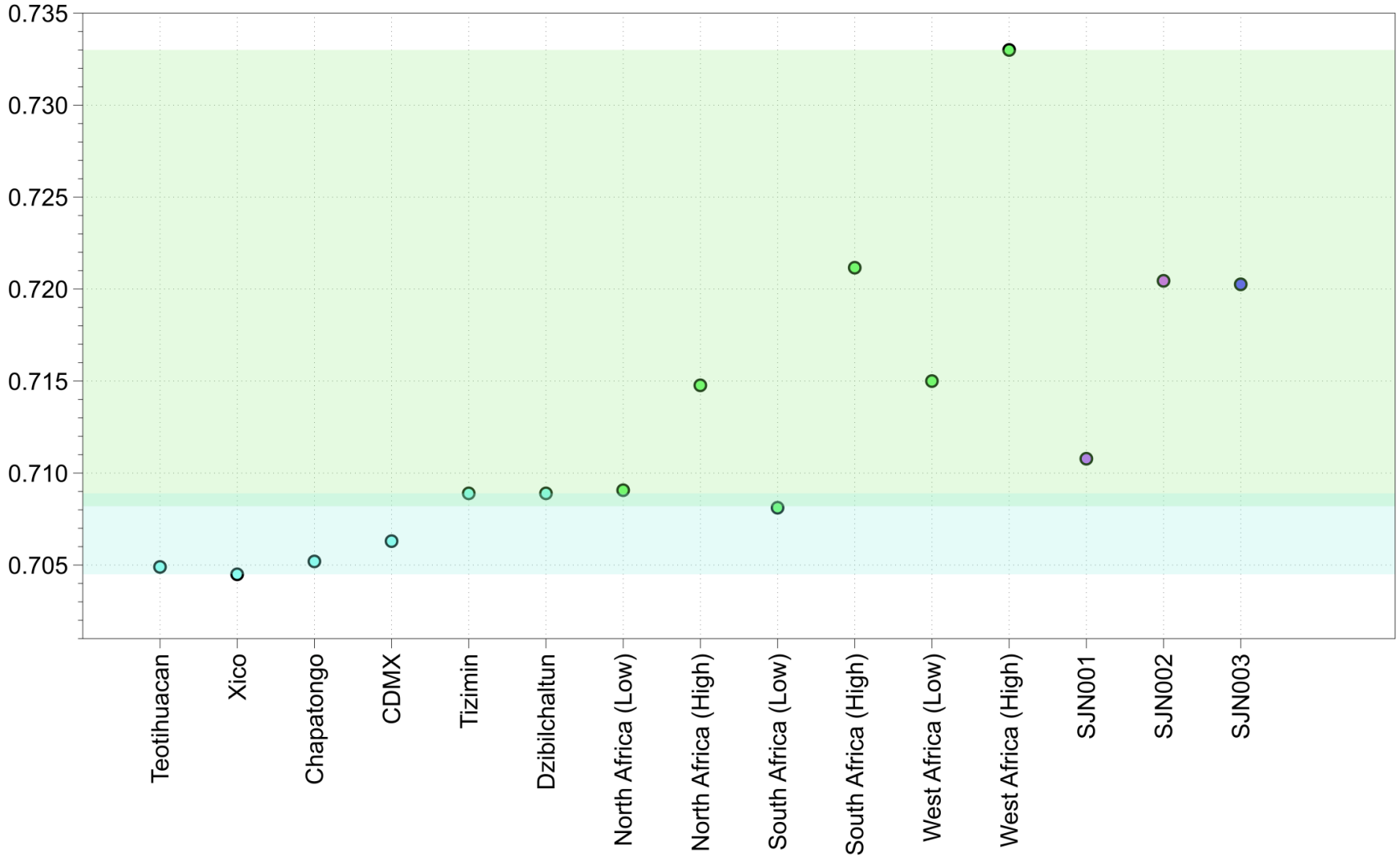


Figure S3. D-statistics of the form D (Chimp, SJN00X; Target, Mende), related to STAR Methods, *F3* and D tests. Plotted values correspond to D ± 3 SE. Red marks indicate values statistically different from zero. **A.** D (Chimp, SJN001; X, Mende). **B.** D (Chimp, SJN002; X, Bantu Ovambo). **C.** D (Chimp, SJN003; X, Lemande)



e S4. Typical 87Sr/86Sr ratios for selected places of Mesoamerica and some regions of Africa compared with the three Africans from SJN, related to Figure
urquoise dots and shaded area correspond to the range of $^{87}Sr/^{86}Sr$ ratios in selected Mesoamerican sites in Mexico, including top-ratio places such as Tizimin and chaltun. Green dots and shaded area correspond to the range of 87Sr/86Sr ratios throughout Africa. The SJN individuals correspond to the violet, pink and purple dots.

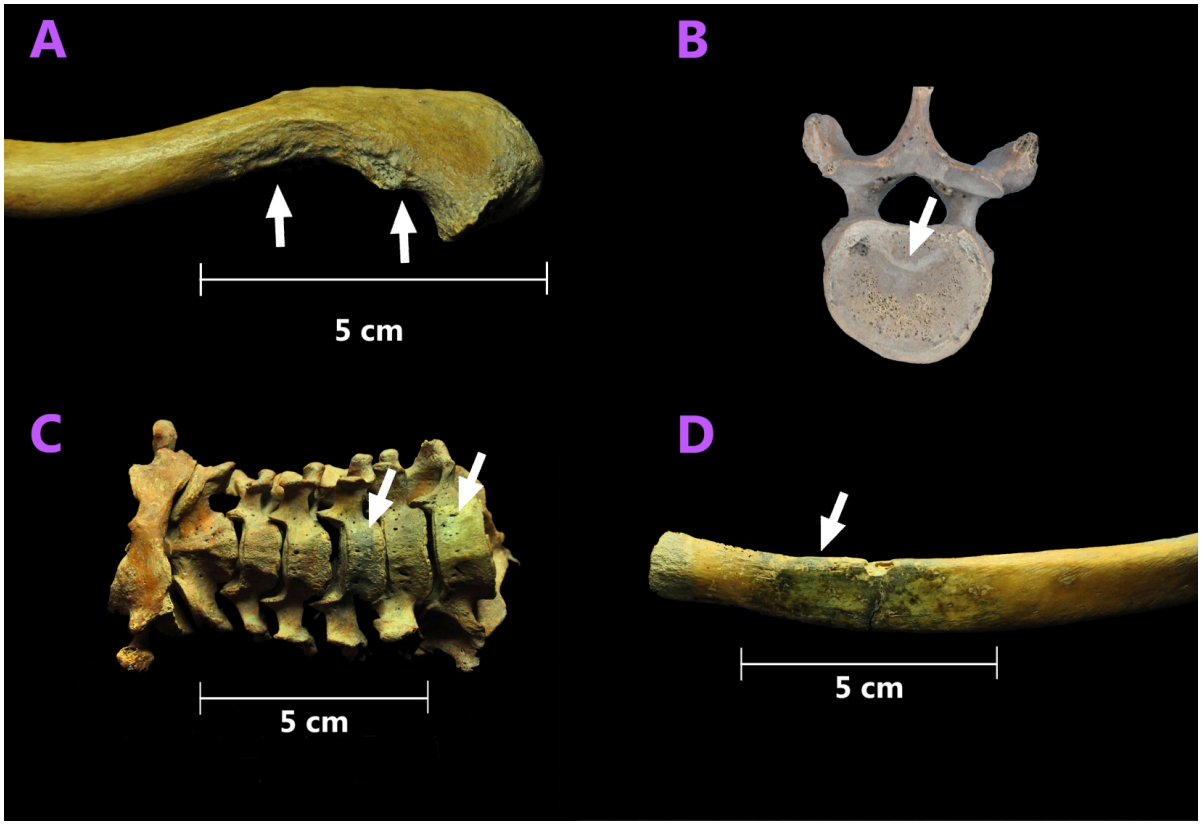


Figure S5: Some osteological findings for individual 150 (SJN001), related to STAR Methods, Osteological assessment, Osteobiography of individual ML8 SL 150 (SJN001). A. Exostosis at the insertion of the coracoclavicular ligament and origin site of the deltoid muscle. **B.** Thoracic vertebra displaying early signs of a developing of Schmörl's hernia on the inferior aspect of the vertebral body. **C.** Green coloration acquired by contact with copper on the cervical vertebrae. **D.** Green coloration acquired by contact with copper on the costal end of a rib diaphysis. Source of the pictures: skeletal collection from San José de los Naturales, Mexico City, in custody of the Laboratory of Osteology, Post Graduate Studies Division, National School of Anthropology and History (ENAH), Mexico.

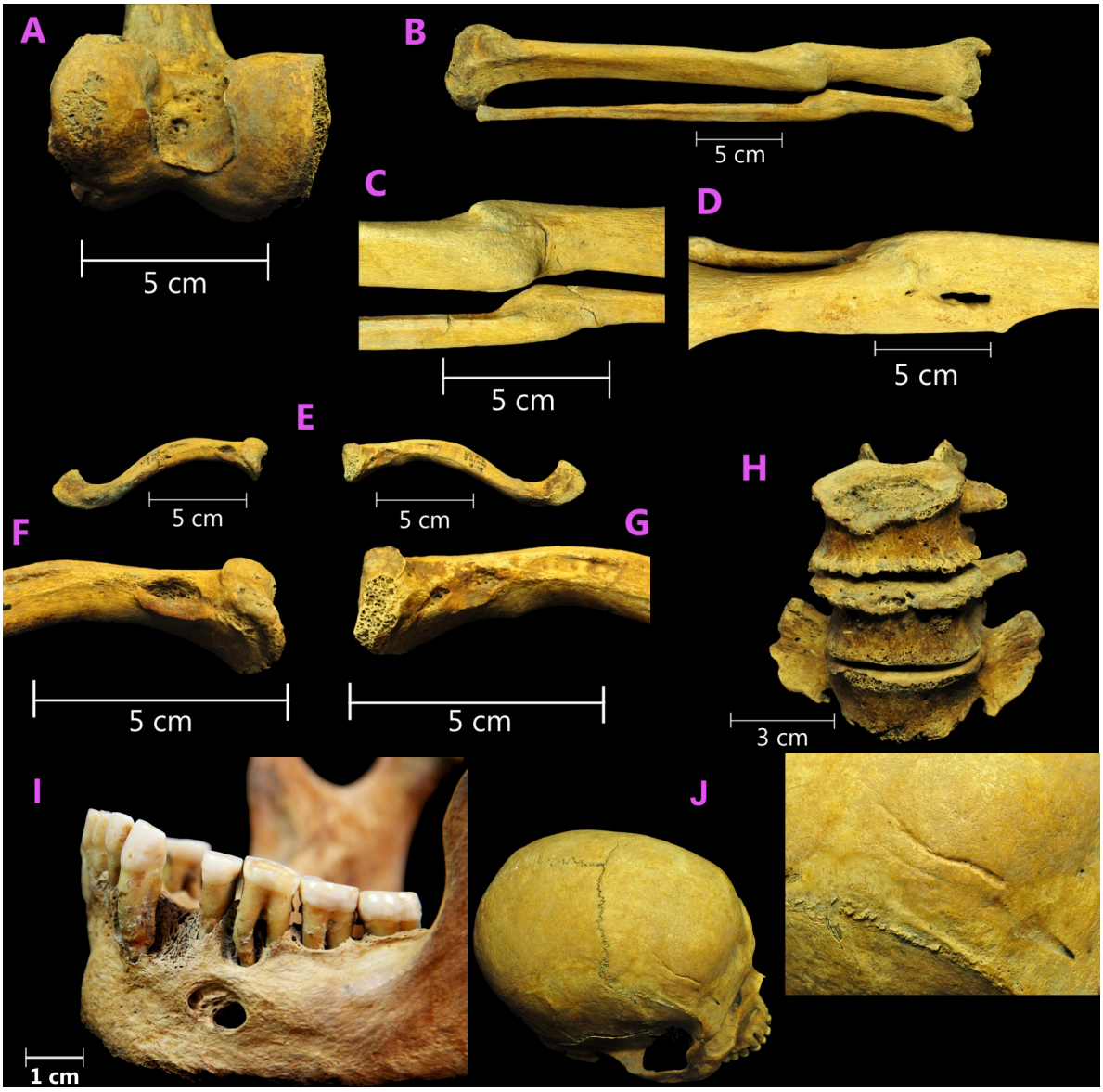


Figure S6: Some osteological findings for individual 214 (SJN002), related to STAR Methods, Osteological assessment, Osteobiography of individual ML8 San José 214 (SJN002). A. The distal surface of the right femur displays osteochondritis dissecans on the medial aspect of the articular surface with some associated lipping on the inferior and posterior edges of the articular surface. **B**. Poorly aligned oblique fracture of the right tibia and fibula resulting in abnormal form through ossification of the fractured skeletal elements. **C.** Close-up of the fracture of the right tibia and fibula. **D.** Medial view of the poorly aligned fracture of the right tibia. **E.** Clavicles from individual 214. **F.** Close-up of left clavicle. **G.** Furrow on the costoclavicular ligament of the right clavicle. **H.** Third, fourth, and fifth lumbar vertebrae are displayed here with excessive osteophytic development on the third and fourth along with signs of compression. **I.** Abscess and skeletal signs of periodontitis on the mandible and cut mark in the frontal bone of the skull. It is possible to appreciate the receding of the mandible and lipping of the alveolar margins. **J.** Cut mark in the frontal bone with signs of bone regeneration. This injury was likely caused by a sharp-pointed object. Source of the pictures: skeletal collection from San José de los Naturales, Mexico City, in custody of the Laboratory of Osteology, Post Graduate Studies Division, National School of Anthropology and History (ENAH), Mexico.

Figure S7: Some osteological findings for individual 296 (SJN003), related to STAR Methods, Osteological assessment, Osteobiography of individual ML8 SLU9B 296 (SJN003). A. Dense sclerotic bone deposition of the right femur likely resulting from well healed periostitis. **B.** Entheseal changes at the muscle insertion of the deltoid of the left clavicle with grained inner surface and raised margins of the bone. Source of the pictures: skeletal collection from San José de los Naturales, Mexico City, in custody of the Laboratory of Osteology, Post Graduate Studies Division, National School of Anthropology and History (ENAH), Mexico.

Reference	Element	$\delta^{13}C$ (in ‰)	$\delta^{15}N$ (in ‰)	C%	N%	C/N	Library	DMG 1st Base 3'	DMG 2nd Base 3'	DMG 1st Base 5'	DMG 2nd Base 5'	Average fragment length	Median fragment length
SJN001	RM_1 (46)	-19.8	13.2	16.6	45.6	3.2	**SJN001.A0102**	0.0678	0.0127	0.0751	0.0124	50.11	50
							SJN001.B0102	0.0309	0.0088	0.0429	0.0071	53.71	51
SJN002	RM^2 (17)	-19.9	10.8	16.6	45.6	3.2	**SJN002.A0102**	0.0710	0.0122	0.0803	0.0119	49.69	49
							SJN002.B0102	0.0776	0.0149	0.0882	0.0147	50.53	50
SJN003	RM_2 (47)	-18.8	10.5	16.6	45.7	3.2	**SJN003.A0102**	0.0474	0.0087	0.0730	0.0106	53.91	51
							SJN003.B0102	0.0443	0.0106	0.0698	0.0114	54.51	51

Table S2. Isotope values of human individual analysed, damage proportions and average and median fragment length for SJN libraries, related to STAR Methods, Stable isotope analysis, diets, and the reservoir effect; Quantification and Statistical Analysis; aDNA Authentication. DMG: Damage.

Position	Reference Call	SJN003	133	Ghana-051	CDC 2575	CDC-1	Gauthier	CDC-2	LMNP-1	HATO	IGU	A10	A12	M2	M3	Fribourg-Blanc	Samoa D	Sei Geringging K403	Kampung Dalan K363	ERS945418	ERS945420	ERS945424	ERS945426	ERS945430	ERS945436	ERS945437	ERS945442	Bosnia A	Iraq B	SS14	Mexico A	94B	Nichols	Sea81-4
46650	G	A	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
105215	G	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
800212	G	A	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
879507	C	T	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1053502	G	A	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1053503	A	G	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
246339	T	C	C	C	C	C	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
376586	C	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	.	.	.	.	.
879200	A	.	N	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	N	G	G	G	G	.	.	.	.	.
1106040	G	.	N	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	.	.	.	.	.
356301	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.
522822	A	.	G	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
624832	T	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
836623	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.
937176	C	.	G	.	.	.	.	.	.	.	N	.	N	.	.	.	.	.	.	.	.	.	.	.	N	.	.	N	N	.	.	N	.	.
1138367	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Table S3: Comparative SNP analysis of *Treponema pallidum* subsp. *pertenue* from Mexico (SJN003, 133) with other *Treponema* spp. genomes, related to STAR Methods, Quantification and Statistical Analysis, Pathogen genomes assemblies, *Treponema pallidum* sub. *pertenue* genome and phylogenetic analyses. Reference call is the allele present in the *Treponema pallidum* sub. *pallidum* Nichols reference genome (GenBank accession number: NC_021490.2). The dot in the cells represents the Reference Call. Strain names corresponding to subspecies *T. pallidum* sub. *pertenue*, *T. pallidum* sub. *endemicum* and *T. pallidum* sub. *pallidum* are highlighted in violet, green and orange respectively. Rows highlighted in yellow correspond to the SNP calls unique to SJN003 or shared with the ex-Convent of Santa Isabel 133 treponemal genome (108) (colonial Mexico *T. pallidum* sub. *pertenue* strains). Rows highlighted in blue show the homoplastic SNP positions common to SJN003 and 133 that are shared with *T. pallidum* sub. *pertenue* genomes. Highlighted in grey is a variant shared between SJN003/133 and strains isolated from patients in Ghana (CDC 2575, CDC-1, Ghana-051). Rows shaded in pink represent SNP positions unique to the 133 treponemal genome.

Supplemental References.

S1. Barbieri, C. et al. (2012) Contrasting maternal and paternal histories in the linguistic context of Burkina Faso. Mol. Biol. Evol. *29*, 1213–1236.
S2. Uren, C. et al. (2016) Fine-Scale Human Population Structure in Southern Africa Reflects Ecogeographic Boundaries. Genetics *204*, 303–14.
S3. Barbieri, C., Butthof, A., Bostoen, K. and Pakendorf, B. (2013) Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia. Eur. J. Hum. Genet. *21*, 430–436.
S4. Torroni, A., Achilli, A., Macaulay, V., Richards, M. and Bandelt, H. J. (2006) Harvesting the fruit of the human mtDNA tree. Trends Genet. *22*, 339–345.
S5. Batini, C. et al. (2011) Insights into the Demographic History of African Pygmies from Complete Mitochondrial Genomes. Mol. Biol. Evol. *28*, 1099–1110.
S6. Oliveira, S. et al. (2018) Matriclans shape populations: Insights from the Angolan Namib Desert into the maternal genetic history of southern Africa. Am. J. Phys. Anthropol. *165*, 518–535.
S7. Barbieri, C. et al. (2014) Migration and interaction in a contact zone: mtDNA variation among Bantu-speakers in southern Africa. PloS One *9*, e99117.
S8. Ingman, M. and Gyllensten, U. (2003) Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. Genome Res. *13*, 1600–1606.
S9. Cerezo, M. et al. (2012) Reconstructing ancient mitochondrial DNA links between Africa and Europe. Genome Res. *22*, 821–826.
S10. Behar, D. M. et al. (2008) The Dawn of Human Matrilineal Diversity. Am. J. Hum. Genet. *82*, 1130–1140.
S11. Brucato, N. et al. (2018) The Comoros Show the Earliest Austronesian Gene Flow into the Swahili Corridor. Am. J. Hum. Genet. *102*, 58–68.
S12. Soares, P. et al. (2012) The Expansion of mtDNA Haplogroup L3 within and out of Africa. Mol. Biol. Evol. *29*, 915–927.
S13. Barbieri, C. et al. (2014) Unraveling the complex maternal history of Southern African Khoisan populations. Am. J. Phys. Anthropol. *153*, 435–448.
S14. Hartmann, A. et al. (2009) Validation of Microarray-Based Resequencing of 93 Worldwide Mitochondrial Genomes. Hum. Mutat. *30*, 115–122.
S15. Gonder, M. K., Mortensen, H. M., Reed, F. A., de Sousa, A. and Tishkoff, S. A. (2007) Whole-mtDNA genome sequence analysis of ancient African lineages. Mol. Biol. Evol. *24*, 757–768.
S16. Schlebusch, C. M. et al. (2017) Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. Science *358*, 652–655.